DOCUMENT RESUME

ED 182 987 FL 010 853

AUTHOR Mackey, William Francis
TITLE La Distance Interlinguistique (Interlinguistic Distance).
INSTITUTION Laval Univ., Quebec (Quebec). International Center for Research on Bilingualism.
PUB DATE 71
NOTE 200p.
LANGUAGE French

EDRS PRICE MF01/PC08 Plus Postage.
DESCRIPTORS Anthropology; Contrastive Linguistics; Diachronic Linguistics; Discourse Analysis; *Language Classification; Language Instruction; *Language Patterns; Language Research; *Language Typology; *Language Universals; *Language Variation; *Linguistic Borrowing; Sociolinguistics; Speech; Synchronic Linguistics; Written Language

ABSTRACT

Standard measures and techniques are needed to establish resemblances and differences among languages. A determination of the degree to which one language or dialect differs from another requires the examination of language universals. To study what languages have in common, attention has to be given to the physical, psychological, and social characteristics common to human beings and the characteristics common to all languages. Universal human characteristics furnish the parameters and the language universals indicate the kinds of measures that are needed. Universal distinctions such as mind and body, or code and speech, lead to a consideration of the measurement of distance among languages from a number of points of view: (1) language and discourse; (2) static and dynamic distance; (3) distance as difference and as conversion; (4) taxonomic and integral distance; (5) semantic and formal distance; (6) direct and indirect distance; and (7) diversity, intensity, and productivity. The utility of each scale and approach to the problem depends on the end in view. The research suggests that a number of problems in pedagogy as well as in linguistics and psychology might be illuminated by the establishment of standard scales of measurement. (AMH)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED1822987

centre international de recherche sur le bilinguisme

publication
B-32

LA DISTANCE INTERLINGUISTIQUE

U S DEPARTMENT OF HEALTH, EDUCATION & WELFARE
NATIONAL INSTITUTE OF EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

CIRB

Gilles Bruyere acting Dir.

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

FL010853

William Francis Mackey

1971

CIRB
ICRB

Centre international de recherches
sur le bilinguisme

International Center for Research
on Bilingualism

LA DISTANCE INTERLINGUISTIQUE

William Francis Mackey

*Centre international de recherches
sur le bilinguisme*

Université Laval, Québec

Les Presses de l'Université Laval
Québec 1971

P R E F A C E

C'est un plaisir pour nous de remercier ici tous tous ceux qui ont contribué à l'élaboration de ce travail.

Tout particulièrement, c'est à Pierre Chantefort et à Jean Darbelnet que nous devons un témoignage de gratitude pour tout ce qu'ils ont fait, tant pour le fond que pour la forme.

Pour le côté mathématique, nous remercions surtout Bernard Morin et Jacques Latouche, qui ont entièrement revu nos calculs.

Enfin, nous tenons à exprimer notre reconnaissance aux secrétaires du CIRB pour la patience qu'elles nous ont témoignée dans la préparation du manuscrit.

William Francis Mackey

CIRB, Université Laval, Québec, 1971

LA DISTANCE INTERLINGUISTIQUE

Table des matières

INTRODUCTION	Page	1
1. Les universaux du langage		5
2. Types de distance interlinguistique		27
3. La distance entre langues		45
4. La distance taxinomique dans le discours		103
5. La distance intégrale dans le discours		139
CONCLUSION		179
BIBLIOGRAPHIE		183

FIGURES ET TABLEAUX

Figure 1	Représentation schématique des vecteurs phonétiques	Page	88
Figure 2	Modèle différentiel		112
Tableau 1	Domaines où l'on pourra rechercher des universaux		11
Tableau 2	Catégories universelles des codes linguistiques		54
Tableau 3	Différences sémantiques du système verbal de l'anglais et du français		57
Tableau 4	Exemples de distance sémantique entre le code lexical du français et celui de l'anglais		63
Tableau 5	Liste des différences formelles entre nominaux		72
Tableau 6	Valeurs prosodiques et notation		85
Tableau 7	Valeurs vectorielles des traits constitutifs des phonèmes		89
Tableau 8	Distance entre les phonèmes de l'anglais et du français		91
Tableau 9	Dimensions de la différence interlinguistique		111
Tableau 10	Signes et symboles pour les mesures taxinomiques		113
Tableau 11	Signes et symboles pour les mesures intégrales		162

INTRODUCTION

On a soutenu que toutes les langues sont des dialectes. Toutefois, ce ne sont pas tous les dialectes qui peuvent prétendre au statut de langue: il faut que ceux qui les parlent soient suffisamment indépendants pour l'affirmer. Le fait que certains dialectes allemands comme le Plattdeutsch et le bavarois ne soient pas mutuellement compréhensibles ne suffit pas à leur conférer le rang de langue et, inversement, ce n'est pas parce qu'il y a inter-compréhension entre les langues scandinaves qu'elles en sont réduites au rang de dialectes. Si l'inter-compréhension est un critère servant à distinguer une langue d'un dialecte, à quel point passera-t-on de l'un à l'autre? Nous ne possédons même pas les outils de mesure nécessaires pour déterminer ce point et si nous les avions, il y aurait des questions plus intéressantes à se poser, comme par exemple: "Quelle est la distance entre langue et dialecte?" Ce n'est là qu'un des aspects de l'intérêt qu'il y a à élaborer des instruments pour mesurer la distance entre langues. Il y en a d'autres pour des domaines tels que l'apprentissage et l'enseignement des langues, le contact des langues, le bilinguisme, les recherches sur les créoles, et les études géographiques et génétiques des langues.

Dans le domaine de l'apprentissage des langues, on accepte généralement l'idée que certaines langues sont plus difficiles que d'autres, en particulier si on se place du point de vue de celui qui les ap-

prend comme langue seconde.

On remarquera, simplement en regardant ou en écoutant des textes dans deux langues, n'importe lesquelles, qu'il y a des différences non seulement dans la façon dont elles se présentent aussi bien par l'écriture que par les sonorités, mais également dans le nombre d'éléments que nous pouvons repérer grâce à la connaissance de notre propre langue. Cependant quand nous nous mettons à apprendre ces langues, les différences se font bien plus apparentes. Le nombre de nouvelles manières de prononcer ou d'écrire variera en fonction de la langue choisie; il en ira de même pour le nombre de différences dans des domaines tels que l'ordre des mots, l'accord ou la longueur des éléments équivalents dans notre propre langue.

Toutefois, le degré de difficulté d'une langue par rapport à une autre reste une question discutable d'appréciation personnelle. On a avancé quelques estimations en notant le temps qu'il faut à certaines personnes pour atteindre un certain degré d'aisance mesuré par certains tests: ainsi, d'après une étude effectuée pour le Corps des volontaires de la paix aux Etats-Unis, il faut presque deux fois plus de temps pour apprendre le chinois que pour apprendre le malais; plus exactement s'il faut à un anglophone éduqué 720 heures pour atteindre un certain degré d'aisance en malais, il lui en faudra 1320 pour atteindre le même degré en chinois.

En ce qui concerne l'enseignement des langues vivantes, il a toujours été délicat de déterminer dans quelle mesure on peut avec pro-

fit transformer certains éléments de la langue maternelle en habitudes exploitables dans la langue étrangère. A quel moment la ressemblance entre langues peut-elle être utile? (Mackey 1972b).

Dans le domaine de l'enseignement bilingue, on renvoie souvent aux minorités ethniques par la langue qu'elles utilisent: espagnol, allemand, français, etc., et on décide parfois que c'est cette langue qui sera leur instrument naturel d'enseignement. Toutefois, il se peut que la variété de langue que parle en fait la minorité soit un mélange de dialecte régional et de mots étrangers bien éloigné de la langue standard internationale à laquelle pensent les responsables de l'enseignement. Quand doit-on décider qu'il vaut la peine de présenter de tels dialectes pour les utilisés comme moyens d'enseignement? En d'autres termes, si on doit enseigner dans une langue autre que celle du pays, est-ce que la variété locale de cette langue est assez proche d'un standard cultivé pour que cela vaille la peine de l'employer? ou bien en est-elle tellement éloignée qu'elle constituera plus un obstacle qu'une aide? Là encore, nous ne savons pas déterminer le degré de différence qui nous permettrait d'étudier sérieusement ce problème.

En ce qui concerne les parents bilingues, dans quelle mesure ce qu'ils parlent est-il influencé par l'une plus que par l'autre langue? Dans le cas d'une langue véritablement mixte et d'une créolisation, dans quelle mesure est-ce que chaque langue est liée à la langue mère? A quel type de langue mixte avons-nous affaire? Quelles sont les proportions du mélange? (Mackey 1970).

Les recherches et les études dans les domaines de la traduction, de l'interprétation et de la stylistique comparée auraient en fin de compte tout à gagner d'une quantification de leurs méthodes de comparaisons de deux langues différentes (Wandruszka 1971).

Dans les études sur les relations entre langues, il est utile de déterminer dans quelle mesure des langues apparentées ou géographiquement voisines ont convergé ou ont divergé, mais il nous faudrait des instruments standards pour mesurer la ressemblance ou la différence (Afendras 1968).

Nous verrons plus loin que beaucoup de ces questions sont liées et que, par exemple, l'inter-compréhension est fonction de la congruence qui à son tour est fonction des interférences bilingues et que toutes deux sont liées au problème de la distance entre langues. A coup sûr, les recherches dans tous ces domaines gagneraient si l'on y introduisait des mesures précises et si on les traitait de façon comparable. C'est dans le but de faire progresser l'économie de ces recherches que nous avons entrepris d'élaborer des techniques de mesures pour comparer les langues.

Le premier problème qui se pose est de déterminer dans quelle mesure une langue ou un dialecte diffère d'un (e) autre. Ceci dépendra de notre connaissance des ressemblances et des différences entre langues, c'est-à-dire des universaux du langage. Cela dépendra également de ce que nous entendons par distance entre langues et de ce que nous considérons être des mesures représentatives. Nous nous occuperons donc d'abord des universaux, des différents types de distance et des techniques de mesure. Commençons par examiner le problème des universaux.

I

UNIVERSAUX DU LANGAGE

1.1 L'homme: pensée, corps, milieu Page 8

1.2 Le langage: langue, discours, communication 10

1.3 Les universaux du langage humain 12

1.3.1 Caractéristiques et contraintes 12

1.3.2 Composantes communes 17

1.4 Les universaux dans la distance interlinguistique 22

I- Les universaux du langage

Dans l'étude des universaux du langage, on retrouve un problème central: montrer que les langues ne sont pas différentes et montrer qu'elles ne peuvent pas différer. Montrer de quelle manière les langues ne diffèrent pas impliquerait que l'on examinât de façon systématique toutes les langues du monde. Montrer, d'un autre côté, de quelle manière les langues ne peuvent différer impliquerait que l'on investigât la nature de l'homme en relation avec le langage et la nature du langage en relation avec l'homme.

Mais comme nous ne sommes pas encore prêts à examiner systématiquement toutes les langues du monde, il nous faudra renoncer à examiner ce que les langues ont en commun pour concentrer notre attention sur ce qu'il est nécessaire que toutes les langues aient en commun. Pour ce faire, nous formulerons les traits nécessairement communs à tous les êtres humains ainsi que les traits que doit comporter toute langue en tant que langue. Nous pourrons ensuite étudier la relation qui existe entre ces universaux et utiliser les résultats pour distinguer différents types de distance entre langues et élaborer des moyens pour les mesurer. Nous allons commencer par examiner les postulats que nous pouvons formuler sur l'homme et son langage.

1.1 L'homme: pensée, corps, milieu

Il semble plausible d'admettre en postulat que toutes les langues humaines ont des traits humains communs. Si pour l'essentiel les hommes sont identiques pour ce qui est du corps et de l'esprit, existe-t-il certaines caractéristiques physiques et mentales liées au langage que l'on puisse qualifier d'universelles?

Il semble que, à propos de l'homme et de son langage, nous puissions poser certaines hypothèses générales:

1- il est capable d'entendre les sons qu'il utilise et de produire les sons qu'il entend;

2- il est capable de grouper les sons pour en faire des énoncés;

3- il est capable d'adresser ces énoncés à différents auditeurs et de les modifier en fonction de l'auditeur;

4- il est capable de lier les sons qu'il entend et qu'il apprend à des différences qu'il peut percevoir dans son environnement;

5- ces caractéristiques des sons et de l'environnement, il les combine en unités fonctionnelles;

6- ces combinaisons s'effectuent pour lui dans la langue qu'il entend et ces regroupements deviennent des conventions qui le tient;

7- il apprend aussi à lier des combinaisons comparables aux opérations de son esprit qui n'ont pas trait à l'environnement: les processus logiques et conceptuels.

8- De cette façon, toutes les activités mentales de l'homme deviennent

liées à ces combinaisons conventionnelles de sons et de concepts: sa perception des séquences sonores et de ce qu'elles représentent, le fait qu'il les conçoive comme entrant dans des catégories de pensée équivalentes, son appréhension de la relation qui existe entre elles, son appréciation des effets produits par le discours sur les personnes qui s'adressent à lui et à qu'il s'adresse, sa façon d'utiliser ces relations et ces appréciations pour fournir des solutions à ses problèmes, le souvenir qu'il a de toutes ces opérations et la façon dont il utilise tout ce qui est ainsi mis en réserve pour prévoir la pensée et l'expression à venir.

S'il est vrai que certains éléments de l'esprit et du corps de l'homme sont identiques où qu'il se trouve et quel qu'il soit, alors il est fort possible que certaines relations entre son activité mentale dans le langage et ses actions physiques dans l'acte de parole soient universelles. Par exemple, il y a certaines limites imposées par la configuration de la pensée et du discours humains; les opérations se déroulent de façon linéaire dans le temps et le discours ne saurait méconnaître les possibilités limitées de l'anatomie humaine, ce qui exclue par exemple des sons qu'il est anatomiquement impossible de produire comme des après-pharyngiennes.

Si par ailleurs, tous les milieux dans lesquels l'homme est capable d'exister ont quelque chose en commun, il est également possible que certains de ces traits soient universels. Simplement pour exister, l'homme doit manger, boire, dormir, exercer une activité, se reproduire et par conséquent, nous sommes en droit de supposer que toutes les langues

doivent avoir des étiquettes ou des catégories conceptuelles qui renvoient à ces activités essentielles; c'est pourquoi nous pouvons dire que certaines caractéristiques de l'esprit, du corps de l'homme et de son milieu sont universelles de par leur nature même. Nous pouvons aussi poser en postulat qu'il existe certaines caractéristiques universelles que toutes les langues de par leur nature même doivent comporter.

1.2 Le langage: langue, discours, communication

Toutes les langues doivent comporter certains éléments relatifs au code, au discours et à la communication. Les codes de toutes les langues doivent être composés de symboles et de procédés (les règles) pour les agencer afin de produire des messages (textes ou discours) utilisables aux fins de la communication. Tous les codes ont en commun certains éléments essentiels: Bien que leurs éléments soient toujours en nombre fini, ils peuvent engendrer des phrases en nombre infini; leurs éléments ne sont pas en eux-mêmes des choses, ce sont des substituts (des symboles) des choses. Ces symboles peuvent représenter des opérations universelles de l'esprit telles que l'identité (être) et la négation (ne--pas), des traits universaux de la condition humaine (l'expérience et l'action), des traits universels de la production du message (la référence à ce qui a été dit) et des traits universels de tous les actes de communication (le locuteur et l'auditeur). Chaque code comprend les symboles dont il a besoin pour fonctionner. Tous les codes ont besoin de symboles, pour (1) les opérations logiques et humaines; (2) les caractéristiques du discours et (3) les variables de la communication. Chaque symbole peut

avoir une ou plusieurs formes exprimées dans les éléments communs à toutes les langues. Chaque type de symbole peut être rattaché aux caractéristiques fondamentales du corps de l'homme, de son esprit et de son milieu. La combinaison des deux ensembles fait apparaître les possibilités offertes par ces relations (Tableau I).

TABLEAU I

Domaines où l'on pourra rechercher des universaux

Le langage / L'homme	Pensée	Corps	Milieu
Langue	1	2	3
Discours	4	5	6
Communication	7	8	9

Nous pouvons, pour chacune de ces cases rechercher les traits universels. Ainsi, quelles seront les caractéristiques universelles de 1- les transformations (les propositions, etc.), 2- l'articulation (les phonèmes, les formes des mots, etc.), 3- la catégorisation (la taille, le nombre, etc.), 4- l'énoncé (l'identification, la substitution, la référence, la voix, etc.), 5- l'enchaînement (l'assimilation, la distribution, etc.), 6- l'agencement (la composition, la liaison des mots, etc.), 7- le style (mode, formes pour s'adresser à quelqu'un, etc.), 8- l'expression prosodique (l'intonation, le rythme, etc.), 9- le registre (le registre formel, l'argot, etc.), la forme (langue écrite, langue parlée)

1.3 Les universaux du langage humain

Qu'y a-t-il d'universel dans les impressions de l'homme et dans la façon dont il les désigne pour produire un énoncé? Y a-t-il quelque chose en commun dans les sources physiques de ces impressions ou dans le processus selon lequel on leur applique des désignations? Existe-t-il des composantes communes applicables à tout acte de parole humain? Il se peut que, pour tout acte de parole humain, les processus qui produisent les caractéristiques et les contraintes de toute langue et les composantes qui leur sont communes à toutes partagent certains traits.

1.3.1 Caractéristiques et contraintes

Il semblerait évident qu'il doit y avoir dans l'esprit de l'homme des symboles pour lui permettre de parler de ce que tous les hommes ont en commun: leur corps, leurs besoins: manger, boire, marcher et dormir. Bien qu'il se puisse bien que ces symboles soient universels, leurs caractéristiques ne le sont pas forcément.

Bien que, par exemple, il soit possible que les parties du corps soient universelles, il n'y a pas d'accord universel sur ce que l'on peut consider comme partie du corps. Ainsi certaines langues slaves ou celtes désignent le pied ou la jambe par un seul mot: noga en serbe; cos en erse, ou bien comme le russe ruka désignent le bras et la main du même mot. Certaines langues n'ont qu'un mot pour désigner les extrémités ainsi en espagnol dedo renvoie aux doigts et aux orteils. Toutefois, comme pour d'autres langues, on peut préciser en utilisant des expressions ou des mots composés: en espagnol, on peut dire los

dedos de los pies. Cependant en coréen, ont dit palkarak (l'orteil) et sonkarak (le doigt) et dans cette langue comme dans d'autres langues, on peut prolonger certaines parties du corps ou bien elles peuvent être des extensions d'une catégorie plus générale; ainsi dans la langue que l'on vient de citer, les orteils, les doigts, les cuillers et d'autres instruments peuvent être désignés comme étant des bâtonnets (kacak) d'un certaom type: palkarak (orteil); sonkarak (doigt); sukkarak (cuiller) chotkarak (baguette). Cela ne paraîtra étrange que s'il s'agit d'usages étrangers, mais l'anglais, le français, l'allemand et d'autres langues européennes utilisent des extensions du même ordre: en anglais, on parlera à propos d'un réveil de "mains" (hands= aiguilles) et de "visage" (face= cadran) du "pied" de la table, de l'"oeil" d'une aiguille (eye-chas); en français on ne parlera pas des "mains" du réveil mais des "aiguilles"; mais dans d'autres contextes, en anglais comme en français on distinguera "l'aiguille" de "l'épingle" (needle ~ pin) alors qu'en allemand le mot"nadel" plus général recouvrira les deux types et que l'on aura recours à des composés pour les distinguer (Nähnadel: aiguille et Stecknadel: épingle).

Un grand nombre (pas tous cependant) de ces processus métaphoriques affectent la grammaire des langues et peuvent se formaliser dans tel ou tel trait grammatical comme la position, la terminaison du mot ou une marque formelle telle qu'un affixe. Il est possible que les distinctions logiques soient formalisées logiquement mais appliquées arbitrairement: en effet, toutes les langues n'appliquent pas les catégories logiques de la même façon. Ainsi, nous pouvons penser que la dis-

tinction entre personne et chose est universelle mais il nous suffit de comparer la grammaire de l'anglais et du français pour nous rendre compte que tel n'est pas le cas: en français, on ne fait pas de distinction formelle entre personne et chose, que ce soit dans le système pronominal (il, elle) ou dans les articles (le, la, un, une) alors qu'en anglais on la fait, pas toujours cependant, pour les pronoms (he, she, it) mais pas pour les articles (the, a). Les universaux de la logique ne sont pas les universaux de la grammaire.

Nous devons faire la distinction entre ce qu'une langue possède et comment elle fonctionne. La plupart des langues semblent posséder des catégories logiques mais elles ne sont pas toujours logiques dans leur emploi. Un concept qui entre dans une langue dans la catégorie des choses peut, dans une autre langue, être classé comme action. Le temps dans une langue peut être traité comme espace dans une autre (il faut longtemps pour parcourir une longue route - A long road takes a long time to cover). En raison de ce caractère arbitraire de l'utilisation par les langues des catégories logiques universelles, il serait bien risqué de deviner les traits grammaticaux communs à toutes les langues.

La différence de catégorisation apparaît comme une contrainte imposée aux conventions de la langue. C'est un peu comme si tous ceux qui parlent une langue avaient accepté en bloc un contrat à prendre ou à laisser. Des contraintes de ce genre sont typiques de toute organisation humaine. Si un pharmacien voulait s'installer en France, en Allemagne, au Canada ou aux Etats-Unis, en France et en Allemagne il pourrait se limiter à la vente de médicaments alors qu'aux Etats-Unis et au Canada, il

lui faudrait aussi vendre des produits de beauté, des revues et une foule d'autres articles. Inversement si un commerçant voulait vendre du savon au détail dans ces pays, il lui faudrait ouvrir une parfumerie en France, une Drogerie où l'on ne demande pas d'ordonnance en Allemagne et une pharmacie ou drugstore aux Etats-Unis et au Canada. Si quelqu'un s'occupait de télégraphe, il aurait affaire au service des postes en Angleterre, en France ou en Allemagne, aux compagnies de chemin de fer au Canada et à une compagnie télégraphique aux Etats-Unis, mais s'il avait besoin de faire des appels téléphoniques interurbains, en France, en Allemagne et en Angleterre, il aurait également affaire au service des postes; mais aux Etats-Unis et dans certaines parties du Canada, il serait en rapport avec des compagnies de services publics. De la même façon, l'interdépendance des catégories linguistiques fonctionne comme élément contraignant non seulement sémantiquement mais également formellement. Il se prolonge aussi jusque dans les mécanismes neurologiques de la parole, fonctionnant sous la forme de réflexes nerveux conjoints: ainsi les anglophones qui prononcent des voyelles faisant appel à un mouvement de l'arrière de la langue sont soumis simultanément, par suite de réflexes qui se conditionnent les uns les autres, à l'obligation d'arrondir les lèvres. Le réflexe les empêche de dissocier les deux, ce qui explique qu'ils ont certaines difficultés à prononcer des sons vocaliques étrangers qui ne lient pas les deux traits, comme par exemple la distinction qui existe en français entre rue et roue (voir ci-dessous). Les réflexes coordonnés sont, à tous les niveaux, les marques de la façon dont une langue donnée combine les traits de sens et de son pour constituer les éléments de son code.

Bien que toutes les langues aient des façons de désigner les noms, les verbes et les qualités, ce ne sont pas les mêmes éléments de l'environnement qu'elles rangent dans chacune de ces différentes catégories, c'est-à-dire que, selon les langues, ce qui est considéré comme objet (le nom), action (le verbe) ou qualité (l'adjectif et l'adverbe) différera. Le type de processus "comme si" affecte tous les niveaux de la langue mais sera appliqué de façon différente d'une langue à l'autre.

Il est possible que lorsqu'on parle de langage, plus on s'élève en abstraction et en généralité, plus on a de chances de trouver des caractéristiques communes. Toutes les langues ont des phrases et peuvent exprimer des propositions. Toutes les langues désignent les objets. La façon dont ces désignations sont assemblées en propositions a pu être déterminée aux premiers temps du langage humain par les dichotomies fondamentales de l'homme à mesure qu'il évoluait pour devenir un utilisateur d'outils. On a avancé que c'est lorsque l'homme a réussi à établir une dichotomie entre la main qui travaille ou qui forge et la main passive ou qui tient quelque chose, qu'il a préparé le terrain pour l'établissement d'autres dichotomies de la pensée et de la parole humaines.

Est-ce que cela lui a également permis d'établir une dichotomie analogue entre les types d'impressions qui sont parvenues à ses sens et leur répartition en catégories stables ou durables (les noms , le sujet) et en catégories actives ou opératoires (les actions, le predicat) et d'étendre ce processus, dans le domaine de la connaissance, d'un

côté, à tous les thèmes, et de l'autre à toutes les remarques portant sur ces thèmes? On a noté que, chez l'enfant, la distinction entre l'action et le sujet de l'action s'établit en gros au moment où il apprend à distinguer ses deux mains. Il semblerait qu'il y ait un parallèle entre le développement physique et le développement linguistique de l'enfant. C'est pourquoi un rapprochement systématique entre les observations d'ordre physique et celles d'ordre linguistique au cours du développement de l'enfant donnerait des renseignements sur l'existence d'universaux dans le langage et sur les éléments communs au code, au discours et à la communication.

1.3.2 Eléments communs

Existe-t-il des éléments communs à tous les codes linguistiques? Existe-t-il aussi des éléments communs dans la partie discours et communication du langage? Dans quelles mesures les codes linguistiques sont-ils semblables? Existe-t-il à l'état latent dans la configuration même du cerveau humain des dispositifs et des structures que l'enfant transforme au fur et à mesure qu'il grandit pour aboutir au système d'une langue particulière? S'il existe une grammaire universelle, y a-t-il quelque chose qui lui corresponde dans le domaine biologique? Il semblerait que la nature du comportement symbolique de l'esprit soit universelle. Dans le langage, il accumule

et fait fonctionner un système de symboles organisés de façon à réaliser productivité et efficacité lors de la communication.

Bien que ces symboles soient liés au milieu que le code de la langue découpe en catégories conceptuelles, certains sont si généraux qu'ils renvoient à des traits communs à tous les milieux, certains sont probablement innés et font partie des mécanismes mêmes du cerveau humain, comme, par exemple, l'impossibilité qu'une chose soit et ne soit pas en même temps et dans les mêmes circonstances, la négation, les distinctions qui sont faites entre choses et qualités, actions et états, lieu et temps, sujet de l'action et objet de l'action, contenant et contenu, et les nombreuses relations possibles entre choses et personnes.

Par exemple, si toutes les langues font la distinction entre action et auteur de l'action ou entre prédicat et sujet, avons-nous le droit d'aller plus loin et de supposer qu'elles font la distinction entre sujets animés et sujets inanimés, entre les modifications affectant le temps (quand) et celles qui affectent l'espace (où), entre le pourquoi et le comment d'une opération? Est-ce que toutes font aussi la distinction entre être, avoir et faire? Dans l'affirmative, nous pourrions aussi plaider pour l'existence de types universels de schémas fondamentaux exprimant des relations telles que: qui est qui; quoi est quoi, qui a fait quoi, où est quoi, quand est quoi, qui a quoi, qui fait quoi, etc. Si, en plus, toutes les langues sont capables de

faire la distinction entre les causes et les résultats d'une action, on pourrait supposer qu'elles expriment aussi des relations du type "qui a fait quoi à qui", qui a fait quoi comment", etc. Si cela est vrai, certains concepts de cas comme l'accusatif, le datif, le locatif, le causatif et le temporel pourraient bien être aussi universels, quelle que soit la manière dont ils sont exprimés.

Peut-être peut-on postuler que toutes les langues disposent de moyens grammaticaux pour distinguer les actions de leurs causes et de leurs effets; en d'autres termes, qu'elles ont des catégories de type verbal et des catégories de type subtantimal et peut-être aussi des catégories exprimant la qualité. La question est de savoir jusqu'où l'on peut aller dans ce genre d'hypothèses sur les universaux des codes linguistiques.

Peut-on dire la même chose des composantes habituelles du discours? Il est possible qu'il y ait des universaux du discours comme la nécessité de renvoyer à quelque chose qui a été dit auparavant et le besoin d'éviter la lourdeur d'une répétition en utilisant à la place une forme plus courte. Là encore, les langues ont des façons différentes de répondre à ce besoin universel. Le français mettra le renvoi à quelque chose d'antérieur dans la même catégorie que les déictiques (L'homme, cet inconnu) là où l'anglais le traitera comme la définition (Man, the Unknown). Toutes les langues ont des façons d'associer leurs catégories grammaticales et leurs catégories conceptuelles

pour produire des messages, des façons de lier les mots en agencements (an: glass of milk / all: Tasse Milch), en combinaisons de mots et de locutions (an: come and go, the sooner the better), pour ordonner action et auteur de l'action, pour exprimer la prédication et la voix (who did it, whom was it done to) et des façons de lier les énoncés ou les propositions selon différents types de relations (bien que, parce que, depuis que, pendant que, si . . . il, <le faire>). En d'autres termes, dans chaque langue, les éléments doivent s'animer mutuellement en se combinant pour donner un discours porteur de sens, ce que l'on peut considérer comme la liaison séquentielle de leurs catégories conceptuelles. C'est cette animation réciproque des éléments qui permet aux codes linguistiques de fonctionner pour produire le discours avec une grande économie, d'avoir des désignations valables pour de nombreuses choses différentes et remplissant des fonctions différentes, éléments qui, par eux-mêmes, sont intrinsèquement ambigus tant qu'ils ne sont pas animés par des contextes verbaux ou situationnels. L'efficacité de ces opérations est montrée par la rareté de l'ambigüité et le petit nombre de cas où les contextes eux-mêmes sont ambigus; et d'ailleurs les

langues ont à leur disposition d'autres moyens d'expression pour dissiper ce genre d'ambiguité.

Si les nécessités universelles de la logique affectent le discours des langues, il en va de même pour les universaux de la communication. Pour que la langue fonctionne comme instrument de communication entre deux individus, on peut soutenir que toutes les langues doivent faire des distinctions entre questions et non-questions, entre l'auditeur et le locuteur, permettre d'exprimer l'attitude de l'un envers l'autre. On peut exprimer l'attitude que l'on a envers la personne à qui on s'adresse par le type ou la longueur de la forme en définissant plusieurs catégories d'individus (par exemple, en français, tu/vous, en allemand Du/Sie). La relation envers l'auditeur peut aussi s'exprimer par le fait que l'on admet et que l'on utilise des symboles communs qui se traduisent par exemple par l'argot ou un jargon technique. Différentes langues peuvent avoir recours à des procédés semblables pour traduire l'attitude du sujet de l'énonciation en face de ce dont il parle, par exemple, s'il parle de quelqu'un qui ne veut pas dépenser ses économies pour acheter une voiture neuve, le fait de choisir des mots comme thrift ou avarice reflete l'attitude du locuteur envers la personne dont il parle. Là encore la façon d'appliquer ces distinctions peut être loin d'être cohérente. Beaucoup de langues emploient la même forme pour désigner la personne dont on parle et la personne à qui l'on parle, en particulier s'il s'agit d'une forme de politesse. Les langues disposent aussi de manières d'exprimer ce que l'on attend de la personne à qui l'on s'adresse (pour obtenir une réponse de sa part, on peut utiliser l'inver-

sion de l'ordre des mots (are you?), un mot particulier (Who...?) ou simplement une modification du ton de la voix (intonation interrogative).

Que peut-on conclure après être redescendu des hauteurs un peu floues de la linguistique conjecturale? Il semble que les langues aient peut-être beaucoup de choses en commun dans ce qu'elles peuvent exprimer mais qu'elles soient susceptibles de différer dans la manière de l'exprimer. Sans analyser véritablement toutes les langues, il y a toutes chances pour que les seuls universaux certains aient leur origine dans la conformation de l'homme: son corps, en particulier les organes de la parole, et dans la nature du discours qu'il produit sous forme d'une suite de symboles au cours de ses activités de communication nécessaires. Quelle peut-être l'utilité de ces universaux pour mesurer la distance entre langues.

1.4 Les universaux dans la distance interlinguistique

Comment peut-on utiliser les universaux de l'esprit, du corps et du milieu et ce qu'il y a de commun dans les codes, le discours et la communication pour mesurer les différences entre langues?

Si nous connaissons la nature exacte de ces universaux, nous pourrons les utiliser comme paramètres. Par exemple, nous connaissons tous les organes du corps humains utilisés pour produire la parole et nous connaissons leurs limites physiques. Ils peuvent servir de constantes pour les mesures phonétiques. Nous connaissons aussi certaines choses que tous les hommes ont en commun: le temps, l'espace, les parties du corps et nous pouvons utiliser le nombre de distinctions faites

à l'intérieur de chaque langue pour mesurer la distance entre langues. Entre ces deux extrêmes de la langue qui touchent au domaine physique du son et à celui du sens, il y a une foule de différences de système et de structure possibles. On ne peut mesurer qu'indirectement les différentes formes d'organisation grammaticale de la langue et l'on ne sera probablement pas capable d'obtenir des mesures exactes tant que l'on ne saura pas quels sont les éléments grammaticaux, comme les parties du discours, qui sont universels. Ne connaissant pas la réponse, on peut soit poser qu'il n'y a pas d'universaux grammaticaux, soit extraire les constantes qui se retrouvent dans les grammaires de toutes les langues après les avoir toutes comparées dans les mêmes termes.

En attendant, ces constantes, portant sur la façon dont les langues peuvent différer mais ne diffèrent pas en fait, peuvent si elles existent constituer une erreur uniforme dans toutes les mesures de distance entre grammaires. Par exemple si les grammaires du français et de l'anglais font toutes deux la distinction entre action et auteur de l'action, ce trait commun, si on le prend en considération, réduira la distance entre les deux grammaires. Cependant, s'il se révèle que cette distinction est universelle, elle n'aura plus de valeur de différenciation.

On peut donc utiliser certaines caractéristiques universelles de l'homme, de la pensée et de la communication pour mesurer la distance entre langues. On peut les décomposer en traits distinctifs un peu comme le signal de télévision est analysé avant d'être synthétisé pour reproduire une image sur un écran. Ainsi, on peut décomposer chaque

phonème en traits sonores distinctifs et de la même façon, on peut considérer que chaque catégorie conceptuelle est composé d'un certain nombre de composantes sémantiques.

Les combinaisons, les séquences et les modifications de ces traits peuvent être permises ou obligatoires à la fois dans ce qui doit être exprimé et dans la façon de l'exprimer. Comme nous l'avons déjà signalé, en français, la combinaison de l'arrondissement labial avec le fait que la voyelle est antérieure est permise (/y/ mais ce n'est pas le cas en anglais où l'arrondissement labial entraîne obligatoirement le recul de la langue (/u/). Les séquences consonnes +/-s/ sont permises en anglais (pillars, plurals) mais impossibles en espagnol (pilares, plurales). Ces contraintes font partie des différences significatives entre langues et dans toute mesure complète de la distance entre langues, on devra en tenir compte, directement ou indirectement.

De quelle façon pourra-t-on alors lier les caractéristiques universelles de l'homme et de son langage aux mesures de la distance entre langues? Les caractéristiques universelles de l'homme fournissent les mesures; les composantes universelles du langage indiquent les types de mesure à faire. Puisque toutes les langues doivent posséder certains traits relatifs au code, au discours et à la communication, c'est de ces traits qu'il faudra tenir compte pour mettre au point les types de mesure. La distance entre codes ne sera pas la même que la distance entre les discours qu'ils produisent; et la distance entre langues pour ce qui est de la communication orale ne sera pas la même que pour la communication écrite du fait que les traits statiques de la page imprimée peuvent

différer de multiples façons de traits dynamiques équivalents du fil de la parole.

Ces distinctions universelles du langage sont liées à leur tout aux caractéristiques universelles de l'homme. Le corps humain est l'instrument et la source de la parole en tant que tel et on peut l'utiliser pour mesurer la distance entre codes linguistiques. Les caractéristiques universelles de l'environnement de l'homme ne sont pertinentes que dans la mesure où les codes linguistiques en rendent compte. Les caractéristiques universelles de l'esprit humain, ses processus symboliques, ses méchanismes pour transformer le code en discours et même sa façon de tolérer les illogismes influeront sur toute tentative pour mesurer la différence entre langues.

Entre ces universaux abstraits de l'esprit humain et les universaux concrets du corps humain, on trouvera toutes les diversités des langues du monde. Les universaux nous permettent d'étudier les diversités en mettant en évidence tous les types possibles de distance entre langues et en servant d'instruments pour les mesures.

II

TYPES DE DISTANCE INTERLINGUISTIQUE

2.1 Distance dans la langue et distance dans le discours Page 30

2.2 Distance statique et distance dynamique 33

2.3 Distance en tant que différence et conversion 35

2.3.1 Distance par différence 36

2.3.2 Distance par conversion 37

2.4 Distance taxinomique et distance intégrale 39

2.5 Distance sémantique et distance formelle 40

2.6 Distance directe et distance indirecte 41

2.7 Diversité, intensité et productivité 42

II- Types de distance interlinguistique

Les distinctions universelles entre esprit et corps, code et discours, parole et écriture, langue maternelle et autres langues, la diversité et l'intensité de la conception qu'a l'homme de son milieu vous obligent à consider la mesure de la distance entre langues d'un certain nombre de points de vue différents.

Ces considérations montrent à l'évidence que l'on peut se placer sous des angles différents pour mesurer la distance entre langues et qu'il y a différents types de distance entre langues. Nous pouvons distinguer entre 1) la distance des deux codes et la distance des discours produits grâce aux deux codes; 2) la distance relative à la communication de messages statiques dans l'espace et la distance que l'on notera dans des énoncés qui se déroulent dans le temps; 3) la distance relative aux différences entre langues par opposition à la distance résultant d'un déplacement de la langue maternelle vers une autre langue; 4) la distance en tant qu'accumulation analytique ou taxinomique de différences de catégories en face de la distance en tant que mesure intégrée globale; 5) la distance en tant que mesure directe des différences par opposition aux mesures indirectes qui font appel à une langue intermédiaire; 6) la distance en tant que la diversité, l'intensité et la productivité des différences comparables entre langues.

Du fait que ces différents types ne s'excluent pas tous mutuellement, on peut mesurer la distance entre deux langues en se plaçant sous plusieurs angles différents comme nous allons le démontrer. Mais avant de mettre au point des techniques pour mesurer la distance entre langues de ces différents points de vue, il nous faut examiner ce qu'impliquent les différents types de distance.

2.1 Distance dans la langue et dans le discours: $\left(A\ X\ ,\ \frac{A}{X}\right)$

Comme nous l'avons déjà vu, toutes les langues disposent de systèmes de symboles que l'on appelle tantôt langue, tantôt code et qui sont utilisés de façon conventionnelle pour produire des messages ou des énoncés, activité qui a reçu un certain nombre de noms dont parole et discours. Mesurer la distance entre langues du premier point de vue ne sera vraisemblablement pas la même chose que le faire du second point de vue.

C'est pourquoi, pour mesurer la distance entre une langue A et une langue X nous pouvons partir soit de la description des deux langues soit d'un échantillonnage des deux langues. En d'autres termes, nous pouvons mesurer la distance (D) 1) entre les codes D(AX) ou 2) entre des messages ou des échantillons de discours $D\left(\frac{A}{X}\right)$

La mesure des différences entre deux codes linguistiques dépend de la précision et de la comparabilité des descriptions utilisées. Si, pour une ou pour les deux langues, les renseignements sont incomplets ou erronnés, les comparaisons et les oppositions que l'on en déduira seront vraisemblablem nt inexactes. Si le type ou la méthode de description

(et de notation) utilisés ne sont pas identiques, des différences dans les techniques de description pourront passer pour des différences entre les langues. Ceci explique pourquoi certaines des études contrastives basées sur les deux codes font appel non pas à des descriptions toute faites mais à de nouvelles descriptions avec en vue la comparaison ou l'opposition des deux langues. C'est encore une raison qui explique que deux descriptions contrastives des mêmes langues ou des mêmes dialectes puissent différer.

Si, cependant, on mesure à partir d'échantillons de discours, il faudra trouver des jeux de textes équivalents avec des versions alternatives; mais même si l'on dispose d'un inventaire de toutes les possibilités, on s'apercevra que mesurer à partir du message n'équivaut pas à mesurer à partir du code: en effet, on n'a pas le droit de supposer que codes différents produisent des textes équivalents qui différeront de la même façon que les codes, ou que le degré de différence entre deux textes équivalents correspond au degré de différence entre les deux codes qui ont engendré les textes. Ceci est particulièrement clair quand on passe d'une région à une autre où l'on parle la même langue. Par exemple, il y a des situations où un Anglais dira "Not at all", un Canadien: "You're welcome" et un Américain: "Uh,Uh". Et pourtant ces trois réponses équivalentes à la même situation, au niveau du discours ne signifient pas que nous avons affaire à trois langues différentes.

Entre deux langues différentes, les différents codes peuvent produire certaines alternatives equivalentes qui sont plus proches que d'autres; ainsi la phrase anglaise The name is Robert peut se rendre

en français de deux façons correctes: Je m'appelle Robert ou le nom est Robert, cette dernière forme étant bien plus proche de l'équivalent anglais. Toutefois, cela ne signifie pas que les deux phrases françaises sont interchangeables; bien au contraire, elles sont liées à des situations qui s'excluent réciproquement, la première s'emploie pour les présentations, la seconde pour les questions. La phrase anglaise s'emploie dans les deux cas et en faire autant en français constituerait donc un exemple d'interférence. C'est pourquoi mesurer les différences entre langues à l'aide de textes équivalents n'est valable que lorsque les deux textes jouent dans le même contexte.

Les ressemblances au niveau du code n'entraînent pas nécessairement des ressemblances au niveau du discours. Mesuré d'après le code, on pourrait dire que le vocabulaire anglais et le vocabulaire français sont très proches comme dans la phrase allemande suivante: "Kann (can) man (one) ein (one) Boot (boat) haben (have)? Mais si l'on mesure d'après l'équivalent anglais de cette phrase, les deux langues ne se ressemblent plus autant:

All : Kann man ein Boot haben? / An : May we rent a boat? Bien que de telles ressemblances entre les codes puissent faciliter la compréhension de l'autre langue, elles peuvent constituer au plan de l'expression une source d'interférences. Les codes de deux langues peuvent être proches l'un de l'autre, tandis que pour chacune, la transformation du code en messages équivalents peut diverger profondément.

Inversement, ce n'est parce que deux messages équivalents

se ressemblent que les codes qui les ont produits se ressemblent de la même façon. Une phrase du genre Maria va a casa peut être aussi bien de l'espagnol que de l'italien, mais si l'on devait comparer les deux codes qui ont engendré cette phrase unique, on se rendrait compte que les deux langues sont différentes: Espagnol (indicatif présent): voy, vas, va, vamos, vais, van; Italien (indicatif présent): vado, vai, va, andramo, andate, vanno.

Espagnol	Italien
Yo voy a casa	Io vado a casa
Maria va a casa	
Nosotros vamos a casa	Noi andiamo a casa
Teresa y Maria van a casa	Teresa e Maria vanno a casa

Ainsi, la distance entre deux langues ne sera vraisemblablement pas la même si on la mesure entre leurs codes ou si on la mesure entre leurs messages. Nous en concluerons qu'aucune des deux mesures n'est par elle-même suffisante, pour donner une vue exacte des choses: les deux sont nécessaires et comme elles sont si différentes, nous devons envisager de mesurer séparément la distance entre les codes et la distance au niveau du discours.

2.2 Distance statique et distance dynamique: D, D

La distance entre langues sera aussi différente selon que l'on mesure les résultats statiques de la communication linguistique ou le processus dynamique de génération des énoncés à partir des codes respectifs.

Pour la mesure statique, on considère la langue comme une réserve d'éléments à apparier et à mesurer. On peut traiter deux langues comme s'il s'agissait de deux paysages dont les traits qui se correspondent reçoivent une attention égale. On peut aussi considérer une langue comme un processus cinétique (D), la production dynamique d'énoncés sur une échelle temporelle sans cesse interrompue. Et l'on peut considérer que le point de cette échelle où les éléments des deux langues sont équivalents est une fonction de leur différence.

Le discours étant essentiellement linéaire, on peut considérer qu'il se déroule dans le temps si bien qu'il n'y a aucun moment où un énoncé peut être appréhendé dans sa totalité. Ainsi, la ressemblance de ce qui précède nous fait prévoir ce qui peut suivre, et en vertu de cela, les éléments se correspondant qui viennent d'abord ont une valeur de ressemblance plus grande que ceux qui suivent; donc, la probabilité pour que l'on assigne son équivalent à un élément dans la séquence temporelle dépend du type et du nombre d'éléments qui sont déjà passés.

Du fait que les séquences de phonèmes se présentent non pas dans l'espace mais dans le temps, la position d'un phonème dans l'une de ces séquences temporelles peut être chargée de sens. Si, au lieu de lire et de traduire les textes, on doit comme les interprètes les écouter, la ressemblance des éléments au début des mots ou des groupes de mots a plus de signification qu'à la fin. L'importance de ce fait se retrouve dans la manière naturelle de former les abréviations orales telles que les "math " ou la "gym" au lieu des "matiques" ou de la "mastique". Le facteur opérant n'est toutefois pas le son, mais la position dans la sé-

quence. Certains systèmes d'ordinateurs sont capables de fonctionner, même quand il s'agit de noms propres, sans avoir besoin de plus de cinq lettres par mot, à condition de commencer à compter à partir de la première lettre. C'est un peu comme si une série d'images devait apparaître à la suite les unes des autres en nous fournissant de plus en plus de renseignements au fur et à mesure que se déroule le film. En ce sens, la langue est essentiellement cinétique et si l'on veut mesurer toutes ses caractérisques, il faut tenir compte de celle-ci.

2.3 Distance en tant que différence et conversion. D(a~X), D(A→X)

On peut considérer la distance entre deux langues comme la somme de leurs différences ou comme le nombre de modifications nécessaires pour passer de l'une à l'autre.

Quand on additionne les différences entre deux langues, on peut les considérer comme des forces opposées ou des secteurs combinables en une mesure de divergence (la distance différentielle ou vectorielle). Ou bien encore, on veut considérer ces différences comme la somme de travail ou le nombre d'opérations nécessaires pour transformer une langue en l'autre langue (distance de conversion). Etant une combinaison de forces, la première distance est non orientée; étant une translation d'opérations, la seconde est uni-directionnelle. Examinons les problèmes qu'implique la mesure de chacune.

2.3.1 Distance par différence D(A~X)

Comme elle est non directionnelle, la mesure de la distance différentielle entre une langue A et une langue X s'obtient en additionnant les différences entre les deux langues si bien que D(A~X) = A - X + X - A. Par exemple, si l'on mesure la différence entre les systèmes de déterminants du français et de l'anglais, le fait qu'à l'article indéfini unique (a) correspondent deux articles en français (un, une) constitue la même valeur de différence que le fait qu'au déterminant personnel anglais (his) correspondent deux formes en français: il y a une différence d'un dans les deux cas. Dire que l'anglais (his, her) = (son, sa) en français et compter zéro comme distance reviendrait à ne pas tenir compte des faits suivants: (her = son, sa) - (1 - 2) et (his = son, sa) - 1 - 2); (son = his, her) - (2 - 1) et (sa = his, her) - (2 - 1) si bien que la différence n'est pas zéro mais, en nombres absolus $|1 - 2| + |1 - 2| + |2 - 1| + |2 - 1| = 4$, c'est à dire une distance vectorielle de 4. La distance vectorielle entre la série des déterminants singuliers anglais et la série correspondante en français sera au total 13, chiffre calculé de la façon suivante:

An :	a	the	this	+	that	my	your	his	+	her				
Fr :	un+une	le+la	ce+cette	ce+cette,	ce+cette	mon+ma	ton, ta, votre	son+sa	son+sa	son+sa				
	1	1	1	2	2	1	1	1	2	2	1			
	2	2	2	1	1	2	2	3	2	1	_	2		
	1	1	1	1	1	1	1	2	1	1	1	=	13	

En d'autres termes, si nous ajoutons les différences entre l'anglais et le français portant sur les distinctions établies dans cette série de dé-

terminants à celles entre le français et l'anglais, nous obtenons un total de 11. Si nous ajoutions toutes les différences portant sur tous les éléments sémantiques et formels de ces deux langues, nous aurions une idée de la distance qui les sépare, mais nous ne connaîtrions pas le nombre de distinctions supplémentaires que devrait établir un Français apprenant l'anglais par rapport à celles que devrait établir un Anglais qui veut apprendre le français. Dans ce cas, le point de départ constituerait une différence du fait que la distance entre A et X n'est pas la même qu'entre X et A. Il nous faudrait répondre à cette question: Quels sont les traits linguistiques qui sont convertis et quels sont-ils à l'arrivée?

2.3.2 Distance par conversion D(A→X)

Il y a en pratique une différence entre les deux cas suivants d'apprentissage d'une langue seconde: partir d'une langue maternelle qui fait un grand nombre de distinctions obligatoires pour essayer d'en acquérir une autre qui n'en fait pas tant, ou, inversement, aller d'une langue simple à une langue plus complexe. Les mesures différentielles ne suffisent pas à permettre de découvrir ce type de différence d'ordre pratique; ce qu'il faut, c'est savoir le nombre de distinctions supplémentaires qu'il faut faire pour passer d'une langue donnée A à une autre langue X; en d'autres termes, nous devons mesurer la distance par conversion.

Pour mesurer la distance par une conversion qui est orientée, il faudra calculer le nombre de distinctions absentes dans l'autre lan-

gue. Reprenons l'exemple des déterminants singuliers en français et en anglais: on constate qu'à chaque déterminant anglais correspondent au moins deux déterminants français: un féminin, un masculin si bien que quand on passe de l'anglais au français, on doit faire à chaque fois une distinction supplémentaire: le genre. Par ailleurs, on trouve pour un déterminant (celui de la 2ème personne du singulier) une distinction supplémentaire d'ordre social: celle qui existe entre les formes familières et la forme de politesse. La distance par conversion est la somme des distinctions qui n'existent pas dans la langue de départ: la différence entre la langue d'arrivée (langue visée ou cible: C) et la langue de départ (langue acquise ou source: S); si bien que S→C = C - S = A - X.

Comparons les deux systèmes suivants:

A (Cible)	F. :	un+une		le+la		ce+cette		ce+cette		mon+ma		ton+ta+votre		son+sa		son+sa
		2		2		2		2		2		3		2		2
		1		1		1		1		1		1		1		1
X (Source)	An.:	a		the		this		that		my		your		his		her
	=	1	+	1	+	1	+	1	+	1	+	2	+	1	+	1	=

D(dét sg) (An → F) = 3

On a affaire à seulement deux types de distinction: l'opposition masculin-féminin qui a la productivité la plus élevée et qui affecte presque tous les déterminants, et les noms et l'opposition (ton, ta/votre) limitée au pronom et au déterminant de la 2ème personne du singulier.

Toutefois si l'on mesure dans l'autre sens, le résultat sera très différent: le francophone apprenant l'anglais ne devra faire que quelques distinctions qui n'existent pas dans son système de déterminants mais qui se trouveront dans des aires sémantiques différentes,

si bien que F → An = An - F:

An - F :

A (Cible)	An	:	a		the		this + that		my		your		his + her	
X (Source)	F	:	un une		le la		ce cette		mon ma		ton ta		son sa	
			1		1		2		1		1		2	
		-	1		1		1		1		1		1	
	An - F :		0	+	0	+	1	+	0	+	0	+	1	= 2

D(dét sg) F → An = 2

Le francophone apprenant l'anglais devra établir deux oppositions qui n'existent pas en français, chacune étant d'un type différent: la première à trait à la proximité, la seconde au sexe; cependant toutes deux sont des éléments de grande intensité en raison de la fréquence d'occurrence, mais du fait qu'elles affectent des petites classes de mots leur productivité dans le code est faible (voir ci-dessous).

2.4 Distance taxinomique et distance intégrale: $\dot{D}$, $\bar{D}$

On peut mesurer la distance entre langues soit en prenant les différences catégorie par catégorie (distance taxinomique) soit en essayant d'en déduire des mesures intégrales de façon à obtenir un tableau d'ensemble.

Avec les mesures taxinomiques, nous avons affaire à des classes, des catégories ou des dimensions; on doit considérer chacune séparément et la mesurer à l'aide de procédures adéquates, propres à la

dimension analysée. Par exemple, les différences entre les relations syntactiques comme l'accord entre l'adjectif et le nom ne seront pas mesurées de la même façon que la différence d'appartenance à une classe pour laquelle dans une langue l'élément sera peut-être un adjectif et dans l'autre, un nom.

Pour la mesure intégrale de la distance entre langues, on ne mesurera les différences observables que des équivalents dans leur ensemble et seulement pour autant qu'ils affectent l'oeil ou l'oreille.

Dans le cas de la distance taxinomique, les résultats des mesures apparaîtront sous la forme d'une suite de chiffres, chaque chiffre représentant la distance à l'intérieur de la catégorie ou de la dimension considérée; dans le second cas, la distance apparaîtra sous la forme d'un seul chiffre intégré.

2.5 Distance sémantique et distance formelle: $D^{(A-X)}$, $D_{(A-X)}$

Les études des langues n'ont pas confondu la description de ce qui est dit et la description de la manière dont c'est dit. Cette opposition a été parfois appelée sémantique-formelle, parfois contenu-expression. Sans nier que tous les traits d'une langue, y compris ceux du sens, soient formalisés dans le code, nous suivrons l'usage traditionnel et, faute de meilleurs termes, nous appellerons "sémantique" ce qui est dit (le signifié), et "formel" la façon de le dire (le signifiant).

Nous allons voir que les différences formelles entre lan-

gues ne sont pas équivalentes aux différences sémantiques. Une rapide étude de quelques faux amis le montre bien. Ainsi la forme location est identique en français et en anglais écrits mais dans le code sémantique français, le mot ne recouvre pas la même aire qu'en anglais car dans cette dernière langue, il serait parfois rendu par rentals, to let etc... Si l'on examinait les équivalents sémantiques en anglais et en français de mots tels que administration, information, relation, transmission, page, machine, pour ne prendre que ceux-là, on trouverait la même opposition entre identité formelle et divergence sémantique.

Par conséquent, il est possible que cette opposition soit importante, car l'agencement des catégories conceptuelles n'est pas nécessairement lié à la relation entre mot et formes. Ainsi, le fait qu'une langue répartisse les mots en masculins et féminins ne signifie pas que cette différence ait quelque chose à voir avec les sens de ces mots.

C'est pourquoi, quand on comparera des codes linguistiques, il faudra conserver cette distinction entre distance sémantique $D^{(A-X)}$ et distance formelle $D_{(A-X)}$. Toutefois, quand on mesure la distance dans le discours, cette distinction n'a plus de raison d'être, car les textes comparables étant équivalents, la composante sémantique devient une constante.

2.6 Distance directe et distance indirecte

La distance entre langues peut être directe ou indirecte. S'il est possible de mesurer directement la distance entre deux langues

ou deux dialectes, il peut sembler sans intérêt de procéder différemment. Pourtant, dans certaines études pour lesquelles la mesure de la distance entre langues est nécessaire, il est possible que les seules différences significatives soient indirectes. Par exemple si nous voulons mesurer la différence du point de vue des normes des deux langues d'une communauté bilingue, nous n'obtiendrons aucun résultat en les comparant l'une à l'autre: il faudra plutôt comparer chacune d'elle avec sa forme standard nationale ou régionale. Ou encore, si on étudie deux langues mixtes qui ont une ascendance commune, c'est par des mesures indirectes que l'on pourra trouver dans quelle mesure chacune diverge ainsi que les résultats des différents types de mélange.

Ainsi, on peut mesurer la distance entre langues directement (D = A - A) en comparant les deux langues ou indirectement en faisant appel à une langue ou à un dialecte intermédiaire (J) comme cela est parfois nécessaire en pratique: D = (A - J) - (X - J).

2.7 Diversité, Intensité et Productivité

Quand on essaie de mesurer la distance entre deux langues, on peut désirer connaître non seulement le degré de différence, mais aussi l'importance des différences à la fois pour celui qui parle les langues et pour le fonctionnement des langues elle-même.

La distance entre deux ensembles de traits comparables appartenant à deux langues différentes peut varier sur trois points: la diversité, l'intensité et la productivité. La diversité a trait au nombre de distinctions qu'une langue établit dans sa grammaire, son vocabu-

laire et sa phonologie; l'intensité est la mesure de l'importance de ces distinctions; la productivité indique le nombre d'aires (situations, structures et catégories) auxquelles s'appliquent ces distinctions. On peut mesurer la distance entre langues en faisant la somme des différences relatives à la diversité, l'intensité et la productivité de leurs éléments comparables.

On peut mesurer les différences de diversité en faisant la soustraction entre le nombre des éléments d'une langue et le nombre des éléments correspondants de l'autre langue. Ainsi, pour l'aire sémantique des pronoms de la troisième personne du singulier, la différence entre le français et l'anglais est de 1:

(he, she, it) - (il, elle) = 3 - 2 = 1 (Mackey 1972a).

On peut mesurer les différences d'intensité en utilisant l'échelle d'importance appropriée, selon que les unités comparables sont des éléments du discours tels que les conjonctions (bien que, puisque, si...) les copules logiques (être, sembler...) ou les variables liées au milieu (vin, neige). On peut trouver les échelles adéquates dans les indices de fréquence, d'inclusion, d'extension, de disponibilité et de familiarité (Mackey 1965; Mackey & Savard 1971; Mackey & Laforge 1972).

La productivité a trait au degré de généralité auquel s'applique en fait la distinction. Affecte-t-elle une vaste classe d'éléments (par exemple toutes les phrases, ou tous les noms) ou bien seulement une petite classe? A l'intérieur de la classe, s'applique-t-elle à tous les éléments ou simplement à quelques uns d'entre eux (par exemple les

pluriels du type foot - feet, goose - geese en anglais)? De même, le français oppose obligatoirement l'action effectuée sur soi-même à l'action non effectuée sur soi-même alors qu'on aura le choix en anglais (F: Je me lave / An: I wash, I wash myself). Cette différence potentielle concernant les oppositions obligatoires peut être multipliée par le nombre d'éléments auxquels elles s'appliquent. On peut aussi considérer les différences de productivité comme le nombre de structures ou d'unités dans lesquelles peut s'insérer un élément. Par exemple, en ce qui concerne la formation du pluriel par mutation vocalique (ex: An: mouse - mice / All: Maus - Mäuse), l'allemand est plus productif que l'anglais qui forme généralement son pluriel avec -s. Sur ce point, l'anglais est plus proche de l'espagnol que de son parent germanique. Par conséquent, on pourra mesurer les différences de productivité en comptant pour chaque langue le nombre d'éléments auxquels s'applique la différence et en calculant le nombre d'énoncés possibles que peuvent produire les oppositions (Mackey 1969).

Pour résumer, lorsque l'on mesure le degré de différence entre langues, on doit faire la distinction entre ce que l'on mesure et le point où on le mesure. Ce que l'on mesure peut être les différences elles-mêmes ou la quantité de travail nécessaire pour transformer une langue en une autre (distance par différence ou par conversion). Le point où l'on effectue les mesures peut nous entraîner à comparer des éléments de code ou bien à analyser des échantillons d'énoncés (distance entre codes ou entre discours). Les techniques de mesure à utiliser dépendront du type de distance que l'on veut mesurer.

III

DISTANCE ENTRE LANGUES

3.1 Différences sémantiques Page 47

3.1.1 Les universaux de sens 47

3.1.2 Mesures sémantiques 49

3.1.2.1 Sens grammatical 50

3.1.2.2 Sens lexical 59

3.2 Différences formelles 64

3.2.1 Les universaux de forme 64

3.2.2 Les mesures formelles 68

3.2.2.1 Forme grammaticale 68

3.2.2.2 Forme lexicale 74

3.2.2.3 Forme phonétique 79

a) distance prosodique 80

b) distance allophonique 86

3.3 Limites des mesures entre codes 97

3- Distance entre langues

Lorsque l'on mesure les différences entre les codes de deux langues, il est utile de distinguer ce qui est sémantique de ce qui est formel du fait que chacun de ces domaines pose des problèmes spécifiques.

3.1 Différences sémantiques

Avant d'essayer de mesurer les différences sémantiques entre langues, il importe de considérer les problèmes d'analyse au niveau du vocabulaire et de la grammaire des langues en relation avec d'éventuels universaux sémantiques.

3.1.1 Les universaux de sens

Les distinctions qu'établit une langue ne se limitant pas aux catégories grammaticales. Les catégories conceptuelles des langues sont si variées et se chevauchent tellement qu'il est difficile de les diviser en significations grammaticales et en significations lexicales. On peut imaginer qu'il est possible de les agencer en échelles selon leur pouvoir d'inclusion pour chacune des zones de l'environnement logique y compris la désignation des choses et des relations, des actions et des états que toutes les langues connaissent probablement. La communication qui est un compromis entre le désir d'inclure et le besoin d'ex-

clure est aussi un point de vue sur l'évolution de tous les codes; en effet tous les codes doivent instaurer une certaine forme de compromis entre la précision, l'efficacité et la redondance. Les codes linguistiques disposent de certaines façons de tenir compte de l'auditeur en faisant appel à des registres et à des schémas intonatifs qui transmettent des implications sémantiques. Chaque langue doit aussi s'adapter aux changements qui se produisent dans l'environnement qui la concerne; pour cela, elle fait varier ses catégories par analogie (An: hands of a clock), par combinaisons (All: Stecknadel) et par agencements (An: as a matter of fact, on account of).

Grâce aux moyens dont elle peut disposer, chaque langue doit être capable de désigner des objets (les parties du corps, la nourriture, le vêtement, l'habitation), des propriétés des objets (nombre, couleur, son, forme, odeur, goût), des actions (manger, dormir, marcher), des relations (dans, sur, à, par, pour), des opérations (identification du moi et d'autrui, questions, réponses).

Egalement, grâce aux procédés que nous venons de mentionner (l'analogie et la combinaison) chaque langue associe, de façon originale et souvent apparemment arbitraire, certaines significations (traits sémantiques) à l'une ou l'autre de ces désignations; si bien que le même ensemble de traits (par exemple le tonnerre) pourra être classé dans une langue dans la catégorie des choses, dans une autre, dans la catégorie des actions et dans une autre encore, dans celle des qualités.

En gardant présent à l'esprit ce fait, la découverte des

traits que toutes les langues possèdent (et non pas doivent posséder) peut fort bien accroître toute mesure de distance entre langues; dans ce cas pouvons-nous continuer à mesurer les distances relatives dans le domaine des catégories logico-grammaticales où l'on a le plus de chances de trouver de tels universaux? Si l'on devait organiser les distinctions sémantiques en ensembles, il y aurait d'abord des chances pour que les ensembles se chevauchent continuellement et ensuite, ce qui dans une langue (A) est assez inclusif peut l'être moins dans l'autre langue (X). Par exemple dans son système pronominal, le français ne fait pas la distinction entre personnes et choses (il, elle) alors que l'anglais la fait (he ± she, it). Mais pour les questions, le français (qui, quoi) fait comme l'anglais (who, what) la distinction; par contre, seul l'anglais la fait dans les subordonnées relatives (An. The person who... the thing which / f. la personne qui,... la chose qui).

Si l'on considère les différences d'oppositions faites par les langues, l'arbitraire avec lequel elles les regroupent et le manque de précision des catégories qu'elles créent, comment peut-on essayer de mesurer les différences sémantiques entre langues?

3.1.2 Mesures sémantiques: $D^{(A-X)}$

Bien que l'on ait déjà fait certaines tentatives pour essayer de mesurer le sens, les mesures sémantiques, même pour une seule et même langue, soulèvent d'énormes difficultés. Pour commencer, on n'est toujours pas d'accord sur le sens du sens. Pour un psychologue, cela peut signifier ce que l'on ressent à propos d'un mot; pour un philosophe,

cela peut signifier ce que l'on pense de ce mot. Les psychologues ont essayé de mesurer des différences de sens affectif en comparant les schémas d'association verbale obtenus en testant différents échantillons d'une population; les réponses ont été organisées de façon à faire ressortir ce que l'on appelle la différentielle sémantique. D'un autre côté, les logiciens se sont efforcés de réduire le sens à des catégories logiques.

Tout en reconnaissant la légitimité et l'utilité de telles études, si l'on veut comparer différents codes, on devra prendre en considération les différences de sens d'ordre linguistique plutôt que celles d'ordre affectif ou logique puis essayer d'élaborer des procédures de mesure. Bien qu'il soit difficile de séparer nettement sens grammatical et sens lexical, il est nécessaire de commencer par admettre cette distinction traditionnelle mais commode, car elle implique des façons différentes d'aborder le problème de la mesure.

3.1.2.1 Sens grammatical

Quand un locuteur combine des catégories sémantiques pour produire un énoncé, il se peut qu'il lui faille tenir compte d'autres traits sémantiques qui peuvent varier d'une langue à l'autre. Ainsi, les concepts auxquels renvoie le mot table coïncideront pratiquement en anglais et en français, mais si l'on veut l'utiliser une seconde fois en employant un substitut (it / elle), l'anglophone devra opérer une distinction sémantique obligatoire pour montrer qu'il s'agit d'une chose, tandis que le francophone ne fera qu'une distinction formelle qui répè-

tera la classification dans les féminins déjà annoncée par l'article (la, une). C'est un peu comme si un voyageur débarquant dans un pays inconnu devait apprendre non seulement que ce pays est habité par différentes catégories de personnes reconnaissables soit (sémantiquement) à ce qu'ils font, soit (formellement) à leur apparence ou à leur façon de s'habiller; mais encore il doit aussi apprendre à qui il faut s'adresser pour obtenir la chose désirée.

Comment peut-on mesurer cette obligation sémantique à laquelle les différentes langues soumettent leurs locutions? Le problème consiste à identifier les oppositions obligatoires qui doivent être faites dans chaque langue. Certaines sont évidentes comme en anglais la distinction entre positif et négatif, d'autres le sont moins comme en anglais la forme progressive ou le subjonctif en français. Il est important de se souvenir que l'on ne traite ici que les traits sémantiques, le sens grammatical en quelque sorte et non pas la forme grammaticale. Cette distinction est importante,car des traits qui paraissent équivalents peuvent être sémantiques dans une langue et formels dans l'autre et s'il en est ainsi, cela accroîtra vraisemblablement la distance entre les langues. Prenons un exemple: le français et l'anglais ont tous deux des déterminants possessifs (An: his, her / Fr: son, sa); mais en anglais, la distinction masculin-féminin est sémantique alors qu'elle est formelle en français, le possessif s'accordant non pas avec le possesseur mais avec la classe formelle (le genre) du mot qu'il qualifie. Son jardin / his garden, her garden; sa maison / his house, her house. En d'autres termes, à la différence de la plupart des langues indo-européennes / sauf

peut-être le persan moderne et l'arménien) l'anglais possède un genre naturel (sémantique) tandis que le français comme la plupart des langues de la même famille, possède un genre formel. Cependant, le genre formel est susceptible de varier d'une langue à l'autre ce qui fait qu'ainsi le soleil est masculin en français et féminin en allemand (die Sonne) ou bien la lune féminin en français et masculin en allemand (der Mond) et cette classification pourra paraître déraisonnable à l'étranger: pourquoi un soldat doit-il devenir féminin en français lorsqu'il monte la garde (la sentinelle)? Pour quelles raisons les jeunes filles et les vieilles femmes sont-elles du neutre en allemand (das Mädchen, das Weib)?

Il est peut-être possible de classer les contraintes sémantiques obligatoires de la grammaire en catégories liées à la fonction grammaticale (l'identification par exemple) et à la prédication. Toutefois, il est possible que les éléments constitutifs de ces catégories restent incomplets, car théoriquement toute langue peut renfermer n'importe quel type d'opposition sémantique obligatoire. C'est ce que signifie la case "autre" du tableau 2. En tant que catégorie de prédication par exemple, "autre" pourrait représenter l'opposition entre verbes d'activité et verbes de mouvement, distinction obligatoire en français mais pas en anglais:

An. At home, he walks / Fr. Chez lui, il marche.

An. He walks home / Fr. Il va chez lui à pied.

Dans les catégories d'identification et de fonction, il existe aussi d'autres possibilités qu'il peut être difficile de prédire à priori. L'identification des partitifs positifs par opposition aux négatifs se trouve

en anglais mais non en français (A_n: some - any / F: de). Une fois qu'on aura établi un inventaire complet de ces oppositions (ce qui n'est pas encore le cas) on pourra utiliser la présence ou l'absence de ces facteurs comme un vecteur pour mesurer la différence entre deux langues (Tableau 2).

TABLEAU 2

Catégories universelles des codes linguistiques

I Catégories de fonction

	affirmation question	acteur action	action but	qualifiant qualifié	narrateur récit	autre
A						
X						

II Catégories d'identification

	expression référence	locuteur auditeur	un beaucoup	défini indéfini	humain animal	masculin féminin	général de politesse	interne externe	contenant contenu	animé inanimé	autre
A											
X											

III Catégories de prédication

	positif négatif	passé non-passé	futur non-futur	perfectif imperfectif	certain possible	simple conditionnel	mouvement absence de mouvement	actif passif	liberté obligation	autre
A										
X										

Utiliser un cadre général des oppositions grammaticales possibles comporte certains désavantages. Tout d'abord, il peut être difficile de s'assurer que le cadre contient bien toutes les oppositions possibles; ensuite certaines langues marquent plusieurs oppositions en même temps si bien que l'on n'est jamais certain du nombre et de la nature des oppositions prévues. Il y a par exemple environ quarante distinctions sémantiques possibles pour le cas, c'est à dire environ quarante façons de relier une chose (ou un concept nominal) à une autre. En latin, le nominatif en recouvre une, l'accusatif deux et le datif quatre.

Au lieu de partir d'une liste de toutes les oppositions sémantiques possibles faites dans les grammaires des langues, il est possible d'adopter une procédure moins ambitieuse qui consistera à comparer les deux systèmes grammaticaux d'un point de vue sémantique. Pour ce faire cependant, il faut bien faire attention à ne pas confondre d'une part forme et fonction, d'autre part contraintes et alternatives. Comme nous l'avons vu, une seule forme peut exprimer diverses relations de type différent la personne, le nombre, le cas, le temps, le mode, la voix. Par exemple, le -o du latin amo (j'aime) exprime la première personne du singulier de l'indicatif présent du verbe amare; à elle seule la forme -atur de amatur (il ou elle est aimé(e)) exprime une distinction de voix, de personne et de temps.

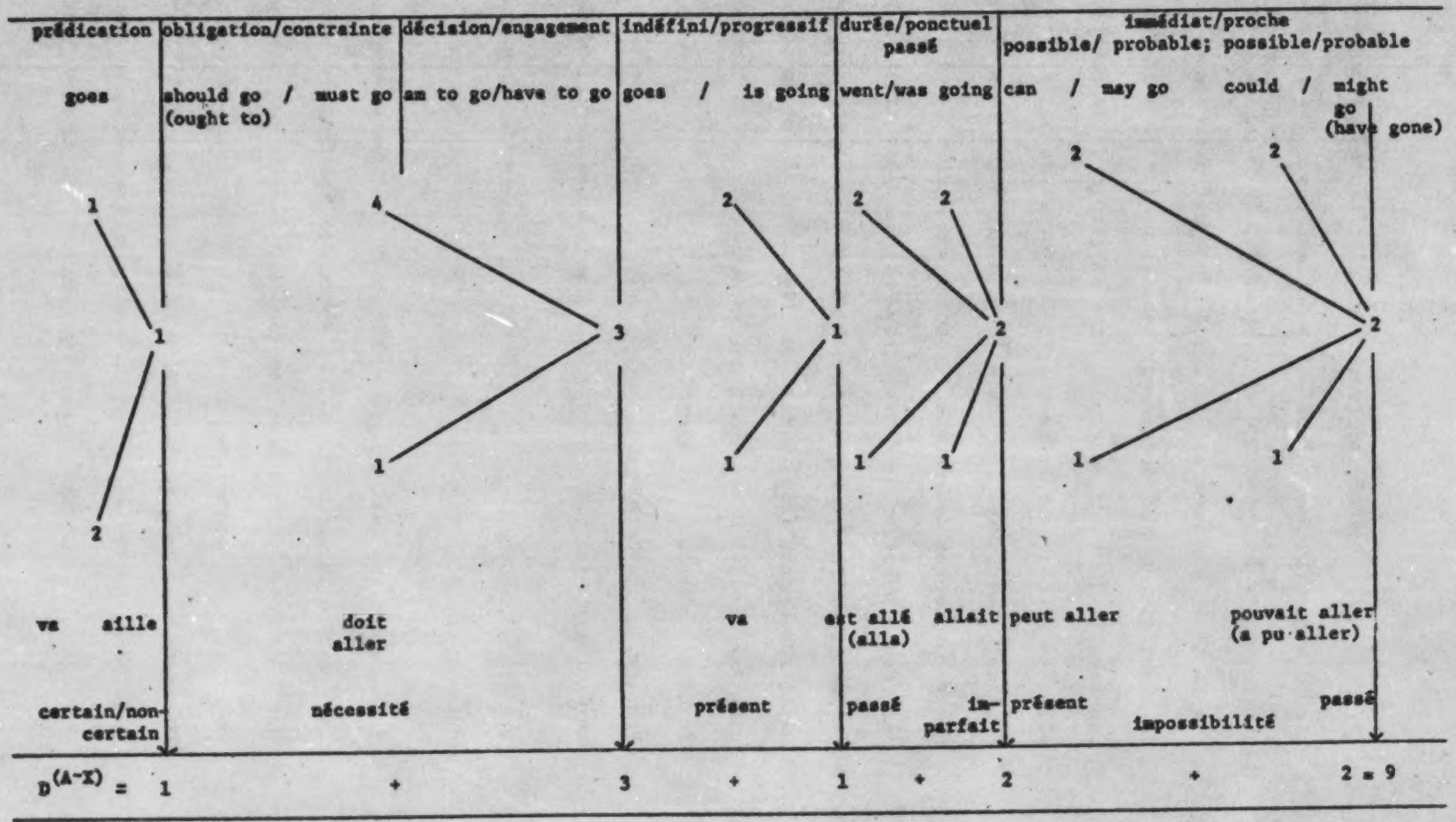
On ne peut traiter une contrainte simplement comme une alternative; prenons les formes verbales suivantes en français: a fini, finit, finissait. Il pourra parfois y avoir trois manières d'envisager le passé mais seulement deux d'entre elles constitueront une contrainte

sémantique, car le francophone peut très bien ne pas tenir compte de la distinction exprimée par le passé simple et s'en tirer parfaitement avec la seule opposition passé composé-imparfait.

C'est pourquoi si nous tenons compte de ces précautions, nous pouvons mettre de côté dans les deux langues les cas où il est nécessaire de faire plus de distinctions sémantiques dans les formes grammaticales d'une langue que dans les formes correspondantes de l'autre langue. Cela revient à la mesure de la distance par conversion de la grammaire des deux langues, quand on mesure dans les deux sens, la différence étant $D^{(A\sim X)} = (A \rightarrow X) + (X \rightarrow A)$. Examinons le système verbal de l'anglais et du français de ce point de vue (Tableau 3).

TABLEAU 3

Différences sémantiques du système verbal de l'anglais et du français



Toutefois, toutes ces différences ne sont pas susceptibles d'avoir la même valeur à l'intérieur des codes respectifs. Sans tenir compte de la fréquence d'utilisation, le degré de contrainte n'est pas le même. On pourrait dire que l'une des différences entre les codes de l'anglais et du français est constituée par le fait que celui-là a conservé le duel indo-européen dans sa catégorie du nombre mais, même si elles se rencontrent fréquemment, les formes où on le trouve sont résiduelles. Toutes les langues doivent avoir des formes résiduelles de ce genre car toutes les langues vivantes évoluent. Ces formes apparaîtront lors des mesures et on devra en évaluer la productivité. La catégorie du nombre en indo-européen comportait bien un duel, par exemple on trouvait en souscrit (tâu) à côté du singulier (sâ) et du pluriel (tè) et bien que ce duel ait disparu en français, il subsiste en anglais comme forme résiduelle. Par exemple:

You can't take both	On ne peut pas prendre les deux
but you can pick the better	mais on peut choisir le meilleur
You can't take all	On ne peut pas tout prendre
but you can pick the best	mais on peut choisir le meilleur

En dernière analyse, la grammaire n'est qu'un moyen dont dispose le locuteur pour établir des relations entre des éléments de son discours porteurs de sens. On ne peut pas toujours décider par oui ou par non ce qui est codé comme grammaire et ce qui ne l'est pas; par exemple, une grammaire de l'anglais variera selon le classement que l'on assignera à des mots comme get et keep. Si on les considère comme have et can, c'est à dire comme ayant des fonctions grammaticales, il nous faudra alors ajouter

deux aspects supplémentaires au système verbal anglais, car ces mots fonctionnels peuvent se combiner avec les formes grammaticales du perfectif (-ed) et de l'imperfectif (-ing) pour produire ce que l'on pourrait appeler le sens grammatical qui, en français, devrait se rendre par des différences de mode, de voix ou de vocabulaire.

	Inchoatif		Duratif	
	Perfectif	Imperfectif	Perfectif	Imperfectif
An	get married	get going	stay married	keep going
Fr	se marient	se mettent en route	restent mariés	continuent

Du fait que beaucoup de catégories grammaticales sémantiques n'existent principalement que dans l'esprit des grammairiens qui établissent des catégories, il sera prudent, quand on mesurera la distance sémantique entre deux langues de ne pas se fier entièrement à eux. Il faut aussi évaluer le nombre de différences qui peuvent être traitées par les catégories conceptuelles qui sont considérées comme faisant partie du vocabulaire.

3.1.2.2 Sens lexical

Comme mesurer n'est pas décrire, les inventaires de traits sémantiques dont nous disposons peuvent être utiles pour établir les indices de différence. On peut trouver beaucoup de ces inventaires, même les abstraits et les plus grammaticaux dans certains dictionnaires unilingues et bilingues, quelque imparfaits qu'ils puissent être, complétés par

les grammaires descriptives et différentielles des deux langues.

Du fait que les regroupements des oppositions d'une langue peuvent chevaucher un certain nombre de regroupements (les catégories) de l'autre langue et inversement, il est nécessaire de mesurer les différences des deux points de vue: (A~X) = (A→X) + (X→A); et du fait que même les dictionnaires et les grammaires les plus complets ne sont pas d'accord sur les catégories de sens et sur ce qu'est l'usage, il faudra tenir compte des deux points de vue quand on se reportera, pour mesurer la distance sémantique, à ces sources particulièrement utiles.

Par exemple, prenons le concept d'homme qui est peut-être universel et examinons ce que deux langues, le français et l'anglais y incluent. On remarque très vite que même si les regroupements, ont beaucoup de points en commun, on trouve des divergences étonnantes. Le concept (homme) ne recouvre pas le même nombre d'aires de sens et à l'intérieur de chaque aire, il ne correspond pas aux mêmes usages spécifiques. Pour chaque concept, on peut utiliser le degré de concordance pour mesurer la distance.

Puisque l'on peut trouver dans un dictionnaire bilingue moderne un inventaire de ce concept et de ses emplois, nous utiliserons le même dictionnaire (par exemple le dictionnaire anglais-français / français/anglais de Harraps édition révisée de 1967) pour mesurer les différences. Si l'on suit les catégories du dictionnaire, on voit qu'il en a établi six pour man (dont une sous catégorie (A = 6) et quatre pour homme (X = 4) ce qui fait un total de dix. Il n'y a que deux catégories

qui soient identiques (:l'humain et le mâle adulte) (A = X) = 2, ce qui donne 20% de chevauchement. Il ne s'agit pas ici des différences stylistiques que peut engendrer le discours (This man wants to know . . . / Ce monsieur désire savoir . . .).

Toutefois, chaque catégorie n'a pas le même nombre d'usages ou de "sens" et à l'intérieur de ceux-ci, il y a aussi des possibilités de chevauchement; c'est pourquoi il vaut mieux compter dans les deux sens, c'est-à-dire les cas ou man = homme et ceux où homme = man. Par exemple, bien que la catégorie mâle adulte (man/homme) existe dans les deux langues, cela ne veut pas dire qu'à l'intérieur les emplois soient toujours équivalents. Ils le sont dans certains cas (I'm your man / Je suis votre homme et Soyez un homme / Be a man), mais pas dans d'autres (Good man! / Bravo et homme à femmes / lady killer). Si nous additionnons ces cas où les langues ne sont pas équivalentes, nous obtenons une autre mesure de la différence.

Pour les six catégories de man (en traduisant chaque emploi de l'anglais par son équivalent en français), nous avons 15 cas d'équivalence (a = x) et 42 cas de non-équivalence (a ≠ x) sur un total de 57 emplois dénombrés. Dans l'autre sens, pour les quatre catégories de homme et en traduisant cette fois du français en anglais (Fr / An) nous trouvons 12 cas d'équivalence (x = a) contre sept cas de non-équivalence (x ≠ a) sur un total de 19. Si nous additionnons toutes ces différences et si nous les exprimons en pourcentage du nombre total de catégories et d'éléments constitutifs, nous arrivons à une distance sémantique de 0,62 en appliquant la formule pondérée suivante:

$$\dot{D}^{(A - x)} = \frac{1}{2}\frac{\Sigma Cat.(A \neq X)}{\Sigma Cat.(A + X)} + \frac{\Sigma\left(a \neq x) = (x \neq a)\right)}{\Sigma\left((a - x) + (x - a)\right)}$$

$$\dot{D}^{(homme-man)} = \frac{1}{2}\left(\frac{6}{10} + \frac{49}{76}\right) = \frac{1}{2}(0,6 + 0,64) = \frac{1,24}{2} = 0,62$$

Cependant, du fait que les dictionnaires et les grammaires sont surchargés de formes, peu productives, rares, à faible fréquence et archaïsantes, il paraît nécessaire de choisir les concepts qui représentent la portion du code véritablement en usage ou les concepts les plus généraux, les plus universaux ou les plus usuels. A titre d'exemple, nous avons pris dix concepts généraux et nous avons abouti à certaines mesures témoin de la distance qui les sépare en français et en anglais (Tableau 4).

Toutefois, l'identité de sens n'indique pas l'identité de forme, pas plus que le degré de différence sémantique n'indique un degré correspondant de différence formelle.

TABLEAU 4

Exemples de distance sémantique entre le code lexical du français et celui de l'anglais

man	/	homme	62,0%
time	/	temps	73,4%
way	/	façon	89,9%
state	/	état	73,3%
world	/	monde	72,7%
life	/	vie	65,9%
come	/	venir	75,4%
see	/	voir	65,4%
high	/	haut	71,6%
house	/	maison	74,6%

Distance sémantique entre 10 éléments lexicaux français et anglais, calculée en pourcentage des différences dans les deux sens portant sur la valence conceptuelle d'un mot anglais et de son équivalent français donné par le dictionnaire.

3.2 Différences formelles

Les traits formels d'une langue sont plus évidents que les significations qu'ils transmettent. Les besoins fondamentaux de la communication et les limitations des organes phonateurs de l'homme concourrent à assigner certaines limites aux traits formels de la langue. Nous commencerons par examiner ces traits formels universels avant d'élaborer des techniques pour en mesurer les différences.

3.2.1 Les universaux de forme

Les besoins de la communication imposent des limites à la longueur d'un énoncé visant à transmettre un trait sémantique conventionnel et plus ce trait est usuel ou fréquent, plus il est susceptible d'être court, mais il y a aussi une limite à la brièveté, car une autre exigence fondamentale de la communication veut que les messages parviennent intacts à destination, (l'esprit de l'auditeur), malgré les distorsions qu'impose la production humaine, qui est de ce fait même imparfaite et les pertes dans la transmission dues au bruit et à la distance, si bien que si un trait se perd, un autre peut le remplacer. En d'autres termes, tout système linguistique doit comporter un certain nombre de traits redondants. Tous les systèmes de toutes les langues naturelles semblent conçus pour produire des redondances, se forment souvent grâce aux contraintes "inutiles" que les langues imposent à ceux qui les parlent. Dans la grammaire d'une langue, elles apparaissent souvent sous la forme de règles qui paraissent inutiles et arbitraires à celui qui apprend cette langue, comme la règle qui veut que l'on emploie à la fois

la préposition et la désinence de la déclinaison pour exprimer la même idée, mais qu'il faut respecter (ex: All: in seinem Beruf - trotz seines Berufs; Lat: a mari - ad mare).

Pour pouvoir produire des énoncés, toutes les langues doivent disposer de certains éléments, comme les mots et les sons, dont on peut dire qu'ils font partie du code de la langue. Pour devenir des instruments de communication, toutes les langues doivent établir certaines distinctions entre ces éléments dans le code et les énoncés qu'elles peuvent produire, sinon une langue ne serait qu'un simple répertoire d'énoncés tout faits qui ne pourraient satisfaire aux besoins complexes de l'homme. C'est pourquoi le nombre d'énoncés potentiels doit être plus grand que le nombre d'éléments du code le la langue; pour que cela soit possible, certains éléments doivent se combiner de différentes façons pour produire des énoncés différents. Ainsi l'énoncé "le chien mord l'homme" ne transmet pas le même message que l'énoncé "l'homme mord le chien". Les différentes façons dont se combinent les éléments sont les structures de la langue, chaque structure renfermant certains types d'éléments dans certaines positions qui constituent les unités de la structure.

Ces unités sont les divisions immédiates de la structure en constituants. Dans certaines langues, ce sont les propositions qui sont les divisions immédiates ou les unités de la structure de la phrase, les syntagmes les unités de la structure de la proposition, les mots les unités de la structure du syntagme etc... De cette façon certaines structures sont plus inclusives que d'autres et ce pouvoir d'inclusion s'ordonne selon une hiérarchie que toutes les langues renferment du fait

que le nombre opérationnel minimum semble être trois par exemple phrase, mot et son.

La différence nécessaire entre code et message implique qu'une structure doit être à même de produire un certain nombre d'énoncés différents en utilisant différentes combinaisons d'unités (le chien mord l'homme, le chien mord l'ours, l'ours mord l'homme, l'ours tue l'homme, etc.); plus il y a d'énoncés différents produits, plus la productivité de la structure est grande. Cela dépend du nombre de constituants de l'unité qui peut occuper la même position, ce qui revient à dire que cela dépend de la profondeur du réservoir dans lequel, en quelque sorte, la structure peut puiser pour produire des combinaisons nouvelles en remplissant les blancs. Par exemple, si la place est occupée par un déterminant (ce, cet, le, la, son, sa, etc.) la profondeur sera moins grande que si c'était un adjectif (bon, mauvais, grand, bleu, etc.) qui occupait la place. La profondeur d'une classe d'unités n'est pas toujours la même d'une langue à l'autre. Enfin la longueur des unités, leur classement et leurs relations varient de langue à langue.

On pourra donc résumer de la façon suivante les caractéristiques formelles universelles des langues.

1) Toutes les langues comprennent des éléments;
2) Ces éléments se composent d'unités et de structures;
3) Certaines sont plus inclusives que d'autres, ce qui donne un caractère essentiellement hiérarchisé à toutes les langues;

4) Les unités apparaissent à l'intérieur des structures dans un certain ordre;
5) Chaque unité a une certaine longueur;
6) Chaque classe d'unité comprend un ou plusieurs membres, ce qui constitue la profondeur de la classe;
7) Chaque unité est d'une certaine façon liée à d'autres unités présentant ainsi un ensemble de relations possibles.

Les éléments formels d'une langue ne sont pas nécessairement déterminés par la façon dont la langue découpe la réalité en catégories, par exemple le fait que deux langues font la distinction entre personnes et choses n'implique pas qu'elles utilisent le même type d'éléments formels pour ce faire. L'une peut opérer cette distinction avec ses déterminants, l'autre à l'aide de non-classes formelles ou de désinences. La puissance sémantique d'une langue, c'est à dire sa capacité d'exprimer et de combiner ses catégories à divers niveaux de signification (logique, grammaticale, écologique, interpersonnelle et intra-textuelle) est une chose et sa puissance formelle est une chose tout à fait différente.

La puissance formelle d'une langue est révélée par les moyens et les choix dont elle dispose pour produire le genre de messages porteurs de sens dont est capable son système sémantique. Ces choix fonctionnent si l'on peut dire à la fois sur un axe vertical et sur un axe horizontal: la structure adéquate est choisie dans la hiérarchie et les unités adéquates se combinent en une séquence linéaire que le locuteur transforme en énoncé. Il est possible que plus on se place haut

dans la hiérarchie moins il y a de différences entre les langues; les choix les plus abstraits sont peut-être universels; mais tout au bas de l'échelle, au niveau de la suite de sons qui constitue l'énoncé, on peut très facilement voir et entendre les différences entre langues: on se trouve à l'autre extrémité du continuum, là où toutes les langues doivent à nouveau déboucher sur le monde physique.

3.2.2 Les mesures formelles

C'est par des éléments comparables que l'on peut mesurer les différences formelles entre codes linguistiques, on peut comparer un mot ou une forme d'une langue avec un mot ou une forme d'une autre langue, on peut opposer un son d'une langue à un son d'une autre langue. Il est donc possible de séparer forme grammaticale, forme lexicale et forme phonétique.

3.2.2.1 Forme grammaticale

Même si les vocabulaires de deux langues se ressemblent beaucoup par la forme et les significations, il se peut que pour les mettre en oeuvre, il faille avoir recours à des activités très différentes dans les deux langues. Bien que la forme police soit identique en français et en anglais et que l'un de ses emplois soit pratiquement le même dans les deux langues, quand l'anglophone veut employer ce mot dans une phrase française, il lui faut tenir compte des différences grammaticales entre les deux langues non seulement de la signification grammaticale qui classe cette notion avec le singulier et non les pluriels mais aussi de

la différence grammaticale formelle qui en fait un féminin (la police) et non un masculin; de plus, il lui faut conserver cette distinction formelle chaque fois qu'il ajoute un adjectif (la police italienne). C'est comme si notre voyageur débarquant en pays étranger voyait des objets connus associés à des personnes inconnues et qu'il lui faille se prononcer sur les personnes, si bien qu'il est forcé de choisir entre des types et des relations dont il ignorait l'existence. Il lui faut simultanément apprendre leur identité, ce qu'ils renferment et ce qu'il peut en faire.

	le			chez elle
police		italien	va	?
	la	italienne		chez lui

Si sa langue n'est pas trop différente, il peut connaître le type d'opposition sans connaître l'opposition elle-même. Par exemple si c'était un Italien et non un Anglais qui ignorait le français, le mot police ne lui semblerait peut-être pas aussi familier du fait qu'en italien l'équivalent est légèrement différent (polizia) mais le fonctionnement du genre formel lui serait familier. Ce qu'il lui faudra apprendre, c'est le genre spécifique des équivalents français et même certains mots qui se ressemblent beaucoup pourront l'induire en erreur. Dans ce cas, il aura trois chances sur quatre de tomber juste et une sur quatre de se tromper car on a estimé à 25% la différence de genre entre mots apparentés français et italiens (Dulsey 1956). De son côté, l'anglophone aura une chance sur deux de se tromper pour le genre des mots apparentés en français, mais il lui faudra d'abord se familiariser avec l'idée de genre formel et

aussi apprendre le genre de chaque mot. Cette considération nous conduit à la notion de niveau ou de hiérarchie lorsque l'on distingue les catégories grammaticales formelles, et elle fait aussi penser qu'il est peu recommandé d'ajouter une opposition à un certain niveau à une autre opposition située à un autre niveau avant que l'échelle des niveaux n'ait été normalisée.

Si nous partons des oppositions formelles les plus générales, comme la distinction entre action-auteur de l'action et qu'en établissant continuellement des sous-classes nous descendions jusqu'au degré de différenciation au niveau de la forme des mots, nous pourrons alors maintenir les niveaux bien distincts et ensuite faire la moyenne des différences.

Si nous procédions ainsi, voici à peu près ce que cela donnerait: y a-t-il des distinctions formelles entre auteur de l'action et action, action et but? Quelles sont-elles? Dans la partie de l'énoncé où l'auteur de l'action intervient, quelles sont les distinctions formelles qui existent? A l'intérieur de chacune de celles-ci (par exemple: les noms) quelles distinctions formelles fait-on (par exemple: le genre)? Par quels procédés s'opèrent ces distinctions (par exemple: affixation, mots de relation, ordre des mots)? Quelles sont les relations qui sont les relations qui sont formellement marquées (par exemple: déterminant-non)? Comment sont-elles marquées (par exemple: accord de la forme)? Quelle portée à la relation (par exemple: 24 formes de déterminatif en français contre 12 en anglais). Quel est le degré de relation formelle (par exemple l'élément th- qui se retrouve dans this-that et these-those)?

Examinons comment les réponses à des questions de ce genre pourraient produire des résultats quantifiables (Tableau 5).

TABLEAU 5

Liste des différences formelles entre nominaux

	Auteur de l'action - action	Action - but	Autre	
A	+	+		
X	+	+		
A~X	0	0		0

	Ordre des mots	Mots de relation	Inflexions	Autre	
A	+		+		
X	+	+			
A~X	0	1	1		2

	Noms	Déterminants	Adjectifs	Pronoms	Autre	
A	+	+	+	+		
X	+	+	+	+		
A~X	0	0	0	0		0

	Nombre	Personne	Genre	Cas	Autre	
A	+			+		
X	+					
A~X	0		1	1		2

	Marqueurs	Affixation	Alternance	Redoublement	Changement interne	
A		+	+			
X	+	+				
A~X	1	0	1		1	3

	Accord Déterminant	Adjectif	Autre	
A	+			1
X	+	+		
A~X	0	1		

	Nombre	Personne	Genre	Cas	Autre	
A		+		+		
X	+	+	+	+		
A~X	1	0	1	0		2

	Portée								
A	the	this-that, these-those		my	your	his-her	our	their	
X	le, la, les	ce-cette	ces	mon, ma, mes	ton ta tes vos votre	son sa ses	notre nos	leur leurs	
A~X	2	0	1	2	4	1	1	1	12

TABLEAU 5 (suite)

Concentration:

the	the	the	this	this	that	that	these	those	my	my	my	your	your	your	his	his	his	her	her	her	our	our	your	your	their	their
-	-	-	s	s	s	s	p	p	1s	1s	1s	2s	2s	2s	3s	3s	3s	3s	3s	3s	1p	1p	2p	2p	3p	3p
sm	sf	p	sm	sf	sm	sf	p	p	1sm	1sf	1sp	2sm	2sf	2sp	3sm	3sf	3sp	3sm	3sf	3sp	1ps	1pp	2ps	2pp	3ps	3pp
le	la	les	ce	cette	ce	cette	ces	ces	mon	ma	mes	ton	ta	tes	son	sa	ses	son	sa	ses	notre	nos	votre	vos	leur	leurs
0	0	0	1	1	1	1	1	1	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2
2	2	1	2	2	2	2	1	1	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3
–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–
2 +	2 +	1 +	1 +	1 +	1 +	1 +	1 +	0 +	0 +	1 +	1 +	1 +	1 +	1 +	1 +	1 +	1 +	1 +	1 +	1 +	1 +	1 +	1 +	1 +	1 +	1 = 27

s = singulier p = pluriel m = masculin , f = féminin , 1 = 1ère personne, 2 = 2e personne, 3 = 3e personne

Différenciation:

this these				that those					
			6/9		4/9				
le les	ce ces	cette ces	ce cette	mon mes	ton tes	son ses	notre nos	votre vos	leur leurs
4/5	4/5	4/8	4/7	2/6	2/6	2/6	4/8	4/8	8/9

ma mes	ta tes	sa ses	mon ma	ton ta	son sa
2/5	2/5	2/5	2/5	2/5	2/5

Une fois que l'on a ainsi mesuré à l'intérieur d'une langue les différences entre les formes des groupes de mots fonctionnels, on les mesure ensuite entre les deux langues c'est à dire en prenant les éléments correspondants dans l'autre langue: the en face de le, la, les, their en face de leur etc.

Toutefois, on peut mesurer ces différences au niveau lexical, en même temps que le vocabulaire des langues ce qui donnera la distance entre formes lexicales.

3.2.2.2 Forme lexicale

Le plus manifeste de tous les éléments linguistiques est le mot. Le locuteur qui parle une langue sans se poser de problèmes sera capable de reconnaître certains mots, même s'il n'est pas capable d'en isoler les éléments syntaxiques, morphologiques ou phonétiques. Le locuteur sans instruction sera aussi capable de reconnaître des ressemblances entre les formes des mots de différentes langues (par exemple: all: Mann et an: man). Il le fait non pas seulement d'après le type et le nombre de sons ou de lettres semblables (ainsi, il n'identifiera pas time et emit bien que les deux mots soient du même type et comportent le même nombre de lettres). Il doit également déceler une ressemblance dans la suite, l'organisation ou la configuration des unités, dans l'ordre où elles sont disposées et la place qu'elles occupent (ainsi time et temps se ressemblent apparemment plus que time et emit). On percevra ces ressemblances de configuration en opérant une juxtaposition linéaire d'un constituant d'une langue et de son équivalent dans l'autre langue. C'est

plus une question de séquence que de constituants. Il y a deux types de corresponsance linéaire: la correspondance par position et la correspondance par équivalence. La correspondance par position consiste à placer l'un en face de l'autre les éléments de deux unités, en fonction de la place qu'ils occupent. Par exemple

An c a t

Fr c h a t

La correspondance par équivalence consiste à placer l'un en face de l'autre les éléments de deux unités en fonction de leur ressemblance ou de leur identité, par exemple

An c a t

Fr ch a t

C'est cette dernière disposition qui constitue la mesure adéquate pour comparer le contenu des unités; la première méthode convient pour mesurer les différences de structure (voir ci-dessous). Si l'on utilise un type de correspondance plutôt que l'autre, les résultats des mesures seront vraisemblablement différents, par exemple: l'anglais cotton et l'espagnol algodòn étant des mots ayant la même origine ont un certain degré de ressemblance formelle. Comparons les deux dispositions pour voir laquelle est la mieux adaptée pour mettre ce fait en évidence:

Correspondance par position	Correspondance par équivalence
An c o t t o n	An c o tt o n
Esp a l g o d ò n	Esp a l g o d ò n
(toutes les paires sont différentes)	(trois paires sont identiques)

En d'autres termes, une façon de calculer montre que les mots ont des

formes entièrement différentes, l'autre montre qu'ils diffèrent d'un peu plus de 50%.

Dans le cas de deux langues ayant des systèmes orthographiques très peu phonologiques, une correspondance visuelle maximale de ce type pourrait faire se ressembler certaines formes plus que ne pourrait le penser le locuteur naïf de l'une ou de l'autre langue. Ainsi

Fr h o m m e

An m a n

On évitera ce piège en faisant correspondre les unités sur la base de la syllabe

An c o t t on An m a n

Esp al g o do' n Fr h o m m e

Sur cette base syllabique, on peut faire correspondre le monosyllabe anglais time et le disyllabe espagnol tiempo

t i m e

t ie m p o

Ainsi, on n'identifiera pas le -e de time et le -e de tiempo. Une fois que l'on a ainsi juxtaposée ou mis en correspondance deux formes, il est possible de mesurer la distance formelle qu'il y a entre elles. On peut mesurer à la fois la distance pour les formes parlées et pour les formes écrites. Toutefois, du fait que la juxtaposition des formes parlées implique la juxtaposition de sons équivalents qui peuvent différer à certains égards (ainsi dans tiempo et time le /t-/ ne se prononce pas de façon identique) nous commencerons par des différences que l'on peut voir et

nous laisserons pour plus tard la mesure plus complexe des différences que l'on peut entendre.

Donc, grâce à la juxtaposition linéaire de deux formes graphiques on est en mesure de voir si deux éléments qui se font face sont identiques ou non; s'ils sont identiques, la distance entre les deux éléments est de toute évidence nulle et on comptera donc zéro; s'ils diffèrent, il y a deux possibilités: soit il n'y a pas de lettre en face auquel cas on peut compter une distance de un; soit il y a une lettre en face qui est différente et on comptera alors une distance de deux. Ceci peut se réduire à une formule simple:

1) Faire correspondre les deux formes de façon à avoir une équivalence syllabique maximale;
2) Compter le nombre total d'unités des deux formes (T);
3) Compter les unités identiques communes qui se font face (C);
4) Soustraire ce nombre (C) du total (T) et on aura une mesure de la distance (D).

Si bien que $D_{(A - X)} = \frac{T - C}{T}$

Reprenons notre dernier exemple (An: time; Esp: tiempo), la distance serait de 4, que l'on calculerait de la façon suivante:

t i m e	T (nombre total des unités)	= 10
t ie m p o soit	C (unités communes identiques)	= 6
0 + 1 + 0 + 2 + 1 = 4	Différence	= 4

$$D_{(time\text{-}tiempo)} = \frac{T - C}{T} = \frac{10 - 6}{10} = \frac{4}{10}$$

On peut aussi réduire ces différences en pourcentage ou les faire varier entre zéro et un ce qui serait nécessaire si l'on veut additionner ou comparer des différences. Ainsi nous voyons que la différence de 4 entre time et tiempo sur un total de 10 équivaut à 40% ou à 0,4; pour l'autre concept (homme) la distance formelle entre la forme anglaise (man) et la forme française (homme) mises en correspondance syllabiquement est de 8 sur 8 soit 1 soit 100%.

La réduction de ces chiffres nous permet de combiner la distance formelle et la distance sémantique exprimée également en pourcentage en les faisant apparaître l'une sans l'autre. Pour l'exemple ci-dessus la distance sémantique $D^{(A - X)}$ pour homme - man) était de 0,62. Si bien que nous avons:

$$D \begin{matrix}(\text{homme} \sim \text{man}) \\ (\text{homme} \sim \text{man})\end{matrix} = \frac{0,62}{1}$$

On peut arriver à une mesure encore plus précisee en pondérant chaque ressemblance en correspondance en fonction de sa place dans la séquence linéaire, car on peut faire valoir que c'est linéairement que l'on perçoit les langues mais comme ce problème entraîne des façons de calculer plus complexes, nous ne l'aborderons que plus tard (voir ci-dessous 5.2.2.).

Poursuivons maintenant nos mesures en examinant les différences entre les façons dont les mots doivent être prononcés puisque l'on peut mesurer les différences de forme lexicale soit d'après l'aspect des mots, soit d'après la façon dont ils se prononcent. C'est notre connaissance des différences de forme phonétique entre les codes qui nous éclaire sur ce dernier point.

3.2.2.3 Forme phonétique

Quand nous avons mesuré la distance entre deux codes, nous avons établi une distinction entre mesures formelles et mesures sémantiques, ces dernières pouvant éventuellement renvoyer au monde physique et exprimer des caractéristiques qui ne se limitent pas à la langue. En descendant de plus en plus bas dans la hiérarchie linguistique pour effectuer nos mesures formelles, nous aboutissons également au monde physique, car les sons, comme les significations, ne se limitent pas à la langue.

Les manières de désigner les significations ou les catégories conceptuelles par des séquences de sons de la langue sont limitées par certaines caractéristiques de l'appareil phonateur. En raison du fait que ces sons sont produits linéairement dans le temps, il doit y avoir des limites entre un son et le suivant. En raison de la confirmation de l'appareil respiratoire humain, il doit aussi exister des unités liées à la respiration; mais dans quelle mesure pouvons-nous postuler le caractère général d'autres traits phonétiques? On pourrait peut-être poser en postulat que la syllabe, la consonne et le son vocalique sont universels sans aller jusqu'à avancer que toutes les langues ont des occlusives, des fricatives, des labiales ou des dentales.

Il serait peut-être plus prudent de parler d'universaux biologiques et de dire que tous les êtres humains normaux ont une bouche, une langue et une gorge. Et puisque tous les sons des langues sont produits dans les limites imposées par les capacités de ces organes humains adaptés à la production des sons, on peut utiliser leurs caractéristiques physiques universelles (fermeture de la bouche, position de la langue etc...)

comme instruments pour mesurer la différence entre les langues tout au bas de la hiérarchie des unités linguistiques.

D'après ces traits physiques universels du corps humain on devrait pouvoir mesurer de façon valable les différences au niveau de la production des sons du langage, car nous pouvons présumer que pour l'essentiel, les organes phonatoires sont les mêmes pour tous les humains du monde entier et qu'à la naissance n'importe qui peut apprendre en puissance n'importe quelle langue. En mesurant la position relative de ces organes au cours de la production de chaque son d'une langue, nous pourrons déterminer dans quelle mesure il différera de n'importe quel son de n'importe quelle autre langue. Quels sont ces organes? Ce sont ceux qui changent de place ou de forme pour produire des différences significatives quand on parle: la langue, les lèvres, les mâchoires, les muscles respiratoires et les cordes vocales. Tous ensemble, ces organes produisent deux types d'unités physiquement mesurables: l'un acoustique et l'autre physiologique. Le premier type fournit l'instrument approprié pour mesurer la distance prosodique entre les syllabes, le second pour mesurer la distance allophonique entre les sons du langage.

a) Distance prosodique

Quand on écoute une langue que l'on ne comprend pas, on est d'abord frappé par la voix qui monte et qui descend et par le rythme des sons qui varient en longueur et en intensité. Certaines personnes apprennent à reconnaître des langues qu'elles ne comprennent pas simplement en prêtant attention à ces traits de rythme et d'intonation.

On peut entendre la différence entre les syllables des langues, car le niveau et la direction du ton de la voix varient ainsi que la longueur et la force relatives des sons qui les constituent. On observe aisément ce phénomène en notant la prononciation des noms de lieu et des mots d'emprunt homophones. Le nom Canada ne sonne pas de la même façon en anglais, en français et en hongrois, pas tant parce que les sons sont prononcés avec de légères différences qu'en raison de la force et de la longueur des syllabes

Anglais	Ca	na	da	forte / faible / faible
Français	Ca	na	da	moyenne / moyenne / forte
Hongrois	Ca	na	da	forte / moyenne / moyenne

C'est grâce à des homonymes de ce type que l'on pourra mesurer la distance syllabique (: prosodique) entre les codes phonétiques.

Il existe au moins deux façons de mesurer les différences syllabiques: instrumentalement et taxinomiquement. Un seul instrument comme l'oscillographe cathodique ou le spectographe acoustique peut révéler des changements dans la longueur ou le ton d'une syllabe, mesurant des différences d'une seule vibration par seconde. L'inconvénient des mesures instrumentales de ce type vient de ce que l'on ne peut utiliser des échantillons que d'un nombre limité de personnes, si bien que des différences individuelles infimes dans la voix et la prononciation ont tendance à masquer les différences entre les codes linguistiques. Cependant, les mesures instrumentales peuvent se révéler utiles pour établir des échelles et des catégories de différence relative utilisables.

Si nous nous contentons d'examiner les différences relatives, nous pouvons les classer dans un certain nombre de catégories en utilisant toute une série de traits distinctifs tels que haut - bas, fort - faible, fixe - changeant, établissant par là à l'intérieur de chaque langue des catégories distinctes pour la force, la longueur, la hauteur et la direction du ton syllabique. Pour chacune de ces quatre variables syllabiques on pourra établir une échelle en affectant chaque point d'une valeur numérique (par exemple bref = 0, moyen = 1 et long = 2). En utilisant ces valeurs comme vecteurs, on pourra opérer des soustractions entre elles en justaposant des syllabes de façon à mesurer en nombres absolus la distance prosodique qui les sépare (Tableau 6).

Pour mesurer la distance prosodique, nous tiendrons donc compte de quatre variables syllabiques: la force, la longueur, la direction et la hauteur (on traitera l'enchaînement avec les mesures phonétiques).

La Force: Dans une séquence parlée, il se peut qu'une syllabe paraisse plus forte ou moins forte que celles qui l'entourent. On mesurera cette variable en affectant la plus faible de l'indice zéro, (pour faire en sorte que les variations de force ne dépendent de la longueur car si on ajoute des zéro, le résultat ne change pas). Ainsi, il est tout à fait possible qu'une syllabe soit en même temps plus longue et moins forte qu'une autre. Nous attribuerons la valeur 2 à la syllabe la plus forte.

La Longueur: On mesurera la longueur individuelle d'une syllabe en donnant la valeur (1) à la plus courte de la langue, (3) à la plus longue et (2) aux autres.

La Direction: Quand on compare deux syllabes, on remarque que le ton peut rester égal pendant toute la longueur de la syllabe ou bien qu'il peut changer dans une, deux directions ou plus pendant la durée de l'émission de la syllabe. Si le ton reste égal, nous donnerons (1) comme valeur à la syllabe; s'il change nous donnerons des valeurs en fonction de la direction et de la quantité de changement.

La Hauteur: Quand on écoute deux syllabes, on peut noter que le ton de l'une est plus haut ou plus bas que celui d'une autre. On pourra mesurer la hauteur en comparant les tons syllabe par syllabe pour voir dans quelle mesure ils correspondent. On donnera (0) comme valeur du ton le plus bas et (5) au ton le plus haut (Tableau 6).

Ainsi, si l'on mesure les différences prosodiques pour la prononciation du mot Canada en français et en anglais, on aboutira aux résultats suivants:

	CA		NA		DA	
RYTHME						
Force						
An	2		0		0	
Fr	1		1		2	
Distance vectorielle	1	+	1	+	2	= 4

Longueur

An	2		0		0	
Fr	1		1		1	
Distance vectorielle	1	+	1	+	1	= 3

INTONATION

	CA		NA		DA	
Direction						
An	1		0		0	
Fr	0		0		0	
Distance vectorielle	1	+	0	+	0	= 1
Hauteur						
An	3		1		1	
Fr	2		2		1	
Distance vectorielle	1	+	1	+	0	= 1

Distance vectorielle totale = 10

TABLEAU 1

Valeurs prosodiques et Notations

(S = Syllabe)

RYTHME

(marqué au dessus de la syllabe)

Force	faible	moyenne	forte
Symbole	(non marqué)	$\overset{\text{'}}{S}$	$\overset{\text{''}}{S}$
Valeur	0	1	2

Longueur	brève	moyenne	longue
Symbole	$\overset{\breve{}}{S}$	$\overset{\bar{}}{S}$	$\overset{\bar{}}{S}$
Valeur	0	1	2

INTONATION

(marquée devant la syllabe)

Direction	statique	descendante	montante	montante - descendante		3 changements
				descendante	montante	
+Symbole	$_{\text{'}}S$	$_{\backslash}S$	$_{/}S$	$_{\wedge}S$	$_{\vee}S$	$_{\sim}S$
Valeur	0	1	2	2	4	5

Hauteur	très basse	basse	mi-basse	mi-haute	haute	très haute
*Symbole	$_{\text{'}}S$	$_{\text{'}}S$	$_{\text{'}}S$	$^{\text{'}}S$	$^{\text{'}}S$	$^{\text{'}}S$
Valeur	0	1	2	3	4	5

[+ les symboles de direction (| \ / ∧ ∨ ∼) s'écrivent sous la ligne]

* les symboles de hauteur s'écrivent ainsi: sous la ligne inférieure $\underline{\text{'}}$ (0), à cheval sur la ligne inférieure $\underline{+}$ (1), sur la ligne inférieure $\underline{\perp}$ (2), sous la ligne supérieure $\overline{\text{'}}$ (3), à cheval sur la ligne supérieure $\overline{+}$ (4), au dessus de la ligne supérieure $\overline{\perp}$ (5).

b) Distance allophonique

Au fur et à mesure que l'on s'approche de la base physique du langage humain on a de plus en plus de chances d'effectuer des mesures exactes. Si l'on applique le type de mesure vectorielle ci-dessus à différents sons du langage, on peut attribuer à ces variables physiques des valeurs graduées correspondant aux différences que l'on peut aussi bien voir qu'entendre. On peut les voir directement ou indirectement à l'aide de techniques photographiques comme la palatographie ou la ciné-radiologie.

Comme tous les sons peuvent être décrits d'après l'action des organes qui les produisent, on peut attribuer une valeur sur une échelle à l'importance de cette action. A l'aide de six échelles correspondant à l'action des différents organes phonateurs on pourra décrire tous les sons pertinents du langage (voir figure 1).

On pourra donc mesurer chaque son de chaque langue en fonction 1) de son point d'articulation; 2) du degré de fermeture ou de resserrement de l'organe affecté; 3) du déroulement dans le temps de l'action des organes (deux ou plus) nécessaires à la production du son; 4) de la forme que prennent les autres organes lors de la production du son; 5) de la direction de la colonne d'air passant par la bouche ou le nez sous l'action du voile du palais et 6) de l'action des cordes vocales. On peut attribuer des valeurs sur une échelle graduée à chacune de ces variables (Tableau 7) (Voir aussi: Grimes & Agard 1959).

Pour mesurer la distance entre les codes phonétiques de

deux langues, on justaposera tous leurs phonèmes et à l'aide de l'échelle de valeurs, on mesurera le degré de différence de leurs traits distinctifs. Par exemple, le phonème anglais /æ/ 1) d'avant: 4

1) d'avant: 4 2) entre ouvert et mi-ouvert: 6,5 3) continu: 3

4) non-constrictif: 0 5) oral: 0 6) sonore: 1

| 4 | 6,5 | 3 | 0 | 0 | 1 |

Figure 1

REPRESENTATION SCHEMATIQUE DES VECTEURS PHONETIQUES

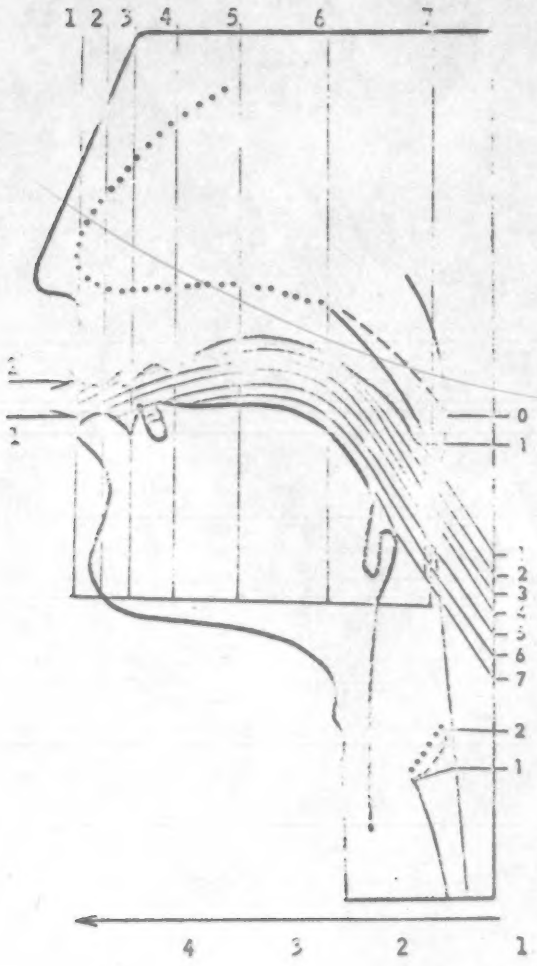


TABLEAU 7

Valeurs vectorielles des traits constitutifs des phonèmes

Valeurs	0	1	2	3	4	5	6	7
Echelles	'	bilabial	labio-dental	interdental	apical	central	d'arrière	glottal
II	'	fermeture	friction	semi-voyelle	fermé	mi-fermé	mi-ouvert	ouvert
III	'	pas de constriction	momentané	continu	allongé			
IV	pas d'articulation	vocalique (détente)	affrication					
V	voile du palais fermé	voile du palais ouvert						
VI	'	vibration des cordes vocales	ouverture des cordes vocales					

TABLEAU 2

Valeurs vectorielles des traits constitutifs des phonèmes

Valeurs	0	1	2	3	4	5	6	7
Echelles I	'	bilabial	labio-dental	interdental	antérieur, apical	central	postérieur	glottal
II	'	fermeture	friction	semi-voyelle	fermé	mi-fermé	mi-ouvert	ouvert
III	'	pas de constriction	momentané	continu	allongé			
IV	pas d'articulation	vocalique (détente)	affrication					
V	voile du palais fermé	voile du palais ouvert						
VI	'	vibration des cordes vocales	ouverture des cordes vocales					

Le phonème français /a/ est: 1) d'avant (4); 2) ouvert (7); 3) continu (3); 4) non-constrictif (0); 5) oral (0); 6) sonore (1)

| 4 | 7 | 8 | 0 | 0 | 1 |

Si on juxtapose les valeurs comparables et que l'on effectue une soustraction en nombres naturels, on obtiendra les distances vectorielles entre les deux sons, ce qui donnera pour le cas présent:

	I	II	III	IV	V	VI	
An /a/	4	6,5	3	0	0	1	
Fr /a/	4	7	3	0	0	1	
D /a - a/	+ 0	+ 0,5	+ 0	+ 0	+ 0	+ 0	= 0,5

La distance vectorielle entre les deux sons est donc de 0,5.

Quand on a de la sorte juxtaposé tous les traits phoniques de deux langues on obtient une table de référence pour les distances vectorielles entre les deux codes phonétiques. A titre d'exemple, nous avons dressé cette table pour mesurer la distance vectorielle entre les sons du français et de l'anglais (Tableaux 8.1, 8.2, 8.3 et 8.4).

Tableau 3.1 An (æ - æ)

Distance entre les phonèmes de l'anglais et du français (sous forme de différence vectorielle entre les traits distinctifs)

Fr (a - o)

							æ	ɑ	ɒ	ɑ̃	ɑɔ	b	d	ʃ	ɛ	ʒ	e	eɪ	f	g	h	ɪ	ɪ	ʤ	k	l	m
							4 6.5 3 0 0 1	5.5 7 3 0 0 1	6 7 3 .5 0 1	4 7 3 1 0 1	4 7 3 1.5 0 1	1 1 3 0 0 1	4.5 1 3 0 0 1	3 2 3 0 0 1	4 5.5 3 0 0 1	5 5 3 0 0 1	5 6 3 0 0 1	4 5.5 3 1 0 1	2 2 3 0 0 2	6 1 3 0 0 1	/ 2 3 1 0 2	4 4 3 .5 0 1	4.5 4.5 3 0 0 1	5 3 3 0 0 1	6 1.5 3 0 0 2	4.5 1 1 .5 0 1	1 1 3 0 1 1
æ	4	7	3	0	0	1	.5	1.5	2.5	1	1.5	9	6.5	6	1.5	3	2	2.5	8	8	10	3.5	3	5	8.5	9	10
ɑ	6	7	3	0	0	1	2.5	.5	.5	3	3.5	11	7.5	8	3.5	3	2	4.5	10	6	8	5.5	4	5	6.5	10	12
ɑ̃	5.5	7	3	0	1	1	3	1	2	3.5	4	11.5	8	8.5	4	3.5	2.5	5	10.5	7.5	9.5	6	4.5	5.5	8	10.5	10.5
b	1	1	3	0	0	1	8.5	10.5	11.5	10	10.5	0	3.5	3	7.5	8	9	8.5	3	5	9	6.5	7	6	6.5	6	1
d	4	1	3	0	0	1	5.5	7.5	8.5	7	7.5	3	.5	2	4.5	5	6	5.5	4	2	6	3.5	4	3	3.5	3	4
e	4	5	3	0	0	1	1.5	3.5	4.5	3	3.5	7	4.5	4	.5	1	2	1.5	6	6	8	1.5	1	3	6.5	7	8
ɛ	4	6	3	0	0	1	.5	2.5	3.5	2	2.5	8	5.5	5	.5	2	1	1.5	7	7	9	2.5	2	4	7.5	8	9
ɛ̃	4	6	3	0	1	1	1.5	3.5	4.5	3	3.5	9	6.5	6	1.5	3	2	2.5	8	8	10	3.5	3	5	8.5	9	8
ə	5	5.5	3	1	0	1	3	3	3	2.5	3	9.5	6	6.5	2	1.5	1.5	1	8.5	6.5	6.5	3	2.5	3.5	7	7.5	10.5
f	2	2	3	0	0	2	7.5	9.5	10.5	9	9.5	3	4.5	2	6.5	7	8	7.5	0	6	6	5.5	6	5	4.5	7	4
g	6	1	3	0	0	1	7.5	6.5	6.5	9	9.5	5	1.5	4	6.5	5	6	7.5	6	0	4	5.5	5	3	1.5	4	6
h	4	4	3	0	0	1	2.5	4.5	5.5	4	4.5	6	3.5	3	1.5	2	3	2.5	5	5	7	.5	1	2	5.5	6	7
ɪ	5	3	3	0	0	1	4.5	4.5	5.5	6	6.5	6	2.5	3	3.5	2	3	4.5	5	3	5	2.5	2	0	3.5	5	7
ʤ	6	1	3	0	0	2	8.5	7.5	7.5	10	10.5	6	2.5	5	7.5	6	7	8.5	5	1	3	6.5	6	4	.5	5	7
l	4	1	1	0	0	1	7.5	9.5	10.5	9	9.5	5	2.5	4	6.5	7	8	7.5	6	4	8	5.5	6	5	5.5	1	6
m	1	1	3	0	1	1	9.5	11.5	12.5	11	11.5	1	4.5	4	8.5	9	10	9.5	4	6	10	7.5	8	7	7.5	7	0
n	4	1	3	0	1	1	6.5	8.5	9.5	8	8.5	4	1.5	3	5.5	6	7	6.5	5	3	7	4.5	5	4	4.5	4	3
ɲ	5	1	3	0	1	1	7.5	7.5	8.5	9	9.5	5	1.5	4	6.5	5	6	7.5	6	2	6	5.5	5	3	3.5	4	4
o	6	5	3	1	0	1	4.5	3.5	2.5	4	4.5	10	6.5	7	3.5	2	3	2.5	9	5	5	3.5	3	4	5.5	8	11
ɔ	5.5	6	3	1	0	1	3	2	2	2.5	3	10.5	7	7.5	3	2.5	1.5	2	9.5	6.5	6.5	4	3.5	4.5	7	8.5	11.5

Tableau 3.2 An (æ - m)

Distance entre les phonèmes de l'anglais et du français (sous forme de différence vectorielle entre les traits distinctifs)

Fr (ɜ - ʒ)

		æ	ɑ	ɒ	ɑɪ	ɑʊ	b	d	ð	ɛ	ɜ	e	eɪ	f	g	h	ɪ	i	j	k	l	m
		4	5.5	6	4	4	1	4.5	3	4	5	5	4	2	6	7	4	4.5	5	6	4.5	1
		6.5	7	7	7	7	1	1	2	5.5	5	6	5.5	2	1	2	4	4.5	3	1.5	1	1
		3	3	.5	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	1	3
		0	0	0	1	1	0	0	0	0	0	0	1	0	0	1	.5	0	0	0	.5	0
		0	1	1	1	1	0	0	0	0	1	0	1	0	0	0	0	1	1	0	0	1
		1	1	1	1	1	1	1	1	1	1	1	1	2	1	2	1	1	1	2	1	1
ɜ	5.5 6 3 1 1 1	4	3	3	3.5	4	11.5	8	8.5	4	3.5	2.5	3	10.5	7.5	7.5	5	4.5	5.5	8	9.5	10.5
ø	4.5 5.5 3 1 0 1	2.5	3.5	3.5	2	2.5	9	5.5	6	1.5	2	2	.5	8	7	7	2.5	2	4	7.5	7	10
œ	4.5 6.5 3 1 0 1	1.5	2.5	2.5	1	1.5	10	6.5	7	2.5	3	2	1.5	9	8	8	3.5	3	5	8.5	8	11
œ̃	4.5 6.5 3 .5 1 1	2	3	3	2.5	3	10.5	7	7.5	3	3.5	2.5	3	9.5	8.5	9.5	4	3.5	5.5	9	8.5	9.5
p	1 1 3 0 0 2	9.5	11.5	12.5	11	11.5	1	4.5	4	8.5	9	10	9.5	2	6	8	7.5	8	7	5.5	7	2
r	4 1 4 0 0 1	6.5	8.5	9.5	8	8.5	4	1.5	3	5.5	6	7	6.5	5	3	7	4.5	5	4	4.5	4	5
n	6.5 1 4 0 0 1	9	8	8	10.5	11	6.5	3	5.5	8	6.5	7.5	9	7.5	1.5	4.5	7	6.5	4.5	3	5.5	7.5
ʊ	6 2 3 0 0 1	6.5	5.5	5.5	8	8.5	6	2.5	3	5.5	4	5	6.5	5	1	3	4.5	4	2	1.5	5	7
s	4 2 3 0 0 2	5.5	7.5	8.5	7	7.5	5	2.5	2	4.5	5	6	5.5	2	4	4	3.5	4	3	2.5	5	6
ʃ	5 2 3 0 0 2	6.5	6.5	7.5	8	8.5	6	2.5	3	5.5	4	5	6.5	3	3	3	4.5	4	2	1.5	5	7
t	4 1 3 0 0 2	6.5	8.5	9.5	8	8.5	4	1.5	3	5.5	6	7	6.5	3	3	5	4.5	5	4	2.5	4	5
u	6 4 3 1 0 1	5.5	4.5	3.5	5	5.5	9	5.5	6	4.5	3	4	3.5	8	4	4	2.5	3	3	4.5	7	10
v	2 2 3 0 0 1	6.5	8.5	9.5	8	8.5	2	3.5	1	5.5	6	7	6.5	1	5	7	4.5	5	4	5.5	6	3
w	1 3 3 0 0 1	6.5	8.5	9.5	8	8.5	2	5.5	3	5.5	6	7	6.5	3	7	9	4.5	5	4	7.5	8	3
y	4.5 4.5 3 1 0 1	3.5	4.5	4.5	3	3.5	8	4.5	5	2.5	2	3	1.5	7	6	6	1.5	1	3	6.5	6	9
ɥ	4 3 3 1 0 1	4.5	6.5	6.5	4	4.5	6	3.5	3	3.5	4	5	2.5	5	5	5	1.5	3	2	5.5	5	7
z	4 2 3 0 0 1	4.5	6.5	7.5	6	6.5	4	1.5	1	3.5	4	5	4.5	3	3	5	2.5	3	2	3.5	4	5
ʒ	5 2 3 0 0 1	5.5	5.5	6.5	7	7.5	5	1.5	2	4.5	3	4	5.5	4	2	4	3.5	3	1	2.5	4	6
-	0 0 0 0 0 0	14.5	16.5	17.5	16	16.5	6	9.5	9	13.5	14	15	14.5	9	11	15	12.5	13	12	12.5	8	7

Tableau 3.3 An (n - dʒ)

Distance entre les phonèmes de l'anglais et du français (sous forme de différence vectorielle entre les traits distinctifs)

Fr (a - o)

	n	ɲ	ɔ	ɔɔ	ɔɪ	ɒ	ɹ	s	ʃ	tʃ	t	θ	u	o	ʌ	v	w	z	ʒ	dʒ	-
	4.5	6	6	6	6	1	5	4	5	5	4.5	3	6	5.5	5.5	2	1	4	5	5	0
	1	1	6	6	6	1.5	3	2	2	2	1.5	2	4	5	6.5	2	3	2	2	2	0
	3	3	3	3	3	3	1	3	3	3	3	3	3	3	3	3	3	3	3	3	0
	0	0	1	1.5	1.5	0	0	0	0	2	0	0	1.5	.5	0	0	0	0	0	2	0
	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	1	1	1	1	1	2	1	2	2	2	2	2	1	1	1	1	1	1	1	1	0
a	4 7 3 0 0 1	7.5	9	4	4.5	4.5	9.5	7	6	7	9	7	7	6.5	4	2	7	7	5	6	8	15
ɑ	6 7 3 0 0 1	8.5	1	2	2.5	2.5	11.5	7	8	7	9	8	9	4.5	3	1	9	9	7	6	8	17
ɑ̃	5.5 7 3 0 1 1	7	6.5	3.5	4	4	12	7.5	8.5	7.5	9.5	8.5	9.5	6	3.5	1.5	9.5	9.5	7.5	6.5	8.5	17.5
b	1 1 3 0 0 1	4.5	6	11	11.5	11.5	1.5	8	5	6	8	5	4	9.5	9	10	2	2	4	5	7	6
d	4 1 3 0 0 1	1.5	3	8	8.5	8.5	4.5	5	2	3	5	2	3	6.5	6	7	3	5	1	2	4	9
e	4 5 3 0 0 1	5.5	7	4	4.5	4.5	7.5	5	4	5	7	5	5	4.5	2	3	5	5	3	4	6	13
ɛ	4 6 3 0 0 1	6.5	8	3	3.5	3.5	8.5	6	5	6	8	6	6	5.5	3	2	6	6	4	5	7	14
ɛ̃	4 6 3 0 1 1	5.5	7	4	4.5	4.5	9.5	7	6	7	9	7	7	6.5	4	3	7	7	5	6	8	15
ə	5 5.5 3 1 0 1	7	7.5	1.5	2	2	10	5.5	6.5	5.5	5.5	6.5	7.5	3	1.5	2.5	7.5	7.5	5.5	4.5	4.5	15.5
f	2 2 3 0 0 2	5.5	7	10	10.5	10.5	1.5	7	2	3	5	3	1	8.5	8	9	1	3	3	4	6	9
g	6 1 3 0 0 1	2.5	1	6	6.5	6.5	6.5	5	4	3	5	3	5	4.5	5	6	5	7	3	2	4	11
l	4 4 3 0 0 1	4.5	6	5	5.5	5.5	6.5	4	3	4	6	4	4	3.5	3	4	4	4	2	3	5	12
j	5 3 3 0 0 1	3.5	4	5	5.5	5.5	6.5	2	3	2	4	3	4	3.5	3	4	4	4	2	1	3	12
k	6 1 3 0 0 2	3.5	2	7	7.5	7.5	5.5	6	3	2	4	2	4	5.5	6	7	6	8	4	3	5	12
l	4 1 1 0 0 1	3.5	5	10	10.5	10.5	6.5	3	4	5	7	4	5	8.5	8	9	5	7	3	4	6	7
m	1 1 3 0 1 1	3.5	5	12	12.5	12.5	2.5	5	6	7	9	6	5	10.5	10	11	3	3	5	6	8	7
n	4 1 3 0 1 1	.5	2	9	9.5	9.5	5.5	6	3	4	6	3	4	7.5	7	8	4	6	2	3	5	10
ɲ	5 1 3 0 1 1	.5	1	8	8.5	8.5	6.5	5	4	3	5	3	5	6.5	6	7	5	7	3	2	4	11
o	6 5 3 1 0 1	7.5	6	1	1.5	1.5	10.5	6	7	6	6	7	8	1.5	1	3	8	8	6	5	5	16
ɔ	5.5 6 3 1 0 1	8	7.5	.5	1	1	11	6.5	7.5	6.5	6.5	7.5	8.5	3	1.5	1.5	8.5	8.5	6.5	5.5	5.5	16.5

Tableau 3.4 An (n - dʒ)

Distance entre les phonèmes de l'anglais et du français (sous forme de différence vectorielle entre les traits distinctifs)

Fr (ɔ̃ - ʒ)

		n	ɲ	ɔ	ɔø	ɔɪ	ŋ	ɹ	s	ʃ	tʃ	t	θ	u	o	ʌ	v	w	z	ʒ	dʒ	-
		4.5	6	6	6	6	1	5	4	5	5	4.5	3	6	5.5	5.5	2	1	4	5	5	0
		1	1	6	6	6	1.5	3	2	2	2	1.5	2	4	5	6.5	2	3	2	2	2	0
		3	3	3	3	3	3	1	3	3	3	3	3	3	3	3	3	3	3	3	3	0
		0	0	1	1.5	1.5	0	0	0	0	2	0	0	1.5	.5	0	0	0	0	0	2	0
		1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
		1	1	1	1	1	2	1	2	2	2	2	2	1	1	1	1	1	1	1	1	0
ɛ̃	5.5 6 3 1 1 1	7	6.5	1.5	2	2	12	7.5	8.5	7.5	7.5	8.5	9.5	4	2.5	2.5	9.5	9.5	7.5	6.5	6.5	17.5
ø	4.5 5.5 3 1 0 1	6.5	8	2	2.5	2.5	9.5	6	6	6	6	6	7	3.5	2	3	7	7	5	5	5	15
œ	4.5 6.5 3 1 0 1	7.5	9	2	2.5	2.5	10.5	7	7	7	7	7	8	4.5	3	2	8	8	6	6	6	16
ɑ̃	4.5 6.5 3 .5 1 1	6	7.5	3.5	4	4	11	7.5	7.5	7.5	8.5	7.5	8.5	6	3.5	2.5	8.5	8.5	6.5	6.5	7.5	16.5
p	1 1 3 0 0 2	5.5	7	12	12.5	12.5	.5	9	4	5	7	4	3	10.5	10	11	3	3	5	6	8	7
r	4 1 4 0 0 1	2.5	4	9	9.5	9.5	5.5	6	3	4	6	3	4	7.5	7	8	4	6	2	3	5	10
ʁ	6.5 1 4 0 0 1	4	2.5	7.5	8	8	8	6.5	5.5	4.5	6.5	4.5	6.5	6	6.5	7.5	6.5	8.5	4.5	3.5	5.5	12.5
ʊ	6 2 3 0 0 1	3.5	2	5	5.5	5.5	6.5	4	3	2	4	3	4	3.5	4	5	4	6	2	1	3	12
s	4 2 3 0 0 2	3.5	5	8	8.5	8.5	3.5	5	0	1	3	1	1	6.5	6	7	3	5	1	2	4	11
ʃ	5 2 3 0 0 2	3.5	4	7	7.5	7.5	4.5	4	1	0	2	1	2	5.5	5	6	4	6	2	1	3	12
t	4 1 3 0 0 2	2.5	4	7	9.5	9.5	3.5	6	1	2	4	1	2	7.5	7	8	4	6	2	3	5	10
u	6 4 3 1 0 1	6.5	5	2	2.5	2.5	9.5	5	6	5	5	6	7	.5	2	4	7	7	5	4	4	15
v	2 2 3 0 0 1	4.5	6	9	9.5	9.5	2.5	6	3	4	6	4	2	7.5	7	8	0	2	2	3	5	8
w	1 3 3 0 0 1	6.5	8	9	9.5	9.5	2.5	6	5	6	8	6	4	7.5	7	8	2	0	4	5	7	8
y	4.5 4.5 3 1 0 1	5.5	7	3	3.5	3.5	8.5	5	5	5	5	5	6	2.5	2	4	6	6	4	4	4	14
ɲ	4 3 3 1 0 1	4.5	6	5	5.5	5.5	6.5	4	3	4	4	4	4	3.5	4	6	4	4	2	3	3	12
z	4 2 3 0 0 1	2.5	4	7	7.5	7.5	4.5	4	1	2	4	2	2	5.5	5	6	2	4	0	1	3	10
ʒ	5 2 3 0 0 1	2.5	3	6	6.5	6.5	5.5	3	2	1	3	2	3	4.5	4	5	3	5	1	0	2	11
-	0 0 0 0 0 0	10.5	12	17	17.5	17.5	7.5	10	11	12	14	11	10	15.5	15	16	8	8	10	11	13	
ɛ̃	0 0 0 1 0 0																					

Du fait qu'il existe rarement une correspondance terme à terme entre les phonèmes de deux langues différentes, il se peut qu'on doive comparer un phonème de la langue A avec deux phonèmes ou plus de la langue X. Par exemple, on pourra comparer /β/ en espagnol avec /b/, /v/ et /w/ en anglais. Quand on compare deux formes grammaticales ou lexicales, les phonèmes individuels peuvent différer en fonction de leur position ou de l'entourage comme en espagnol les deux /d/ dans le mot /dedo/.

Ni la représentation graphique ni la représentation orthographique des éléments formels de deux langues ne donneront un tableau fidèle du degré de ressemblance entre deux formes. Cependant une comparaison phonème par phonème d'équivalents que l'on juxtapose et une mesure des différences entre les traits distinctifs révèleront des ressemblances de forme qui pourraient échapper au regard. Par exemple, que la lettre c soit comparée à k, q ou g ou à toute autre lettre, on notera toujours qu'il y a une différence.

On pourra faire ressortir les ressemblances entre des mots écrits dans des systèmes orthographiques différents en transcrivant une représentation graphique des phonèmes de façon à ce que les différences orthographiques ne masquent pas les ressemblances de forme. Comparons ainsi les paires suivantes.

An	c a t s	An	/ k a t s /
All	k a t z en	All	/ k a t s e n/
An	q u a r t e r	An	/ k w ɔ r t a r /
Esp	c u a r t o	Esp	/ k w a r t o /

Le choix de la transcription variera selon que l'on voudra mesurer la distance entre les formes de la langue, en tant que formes ou que l'on voudra rendre compte de certaines variables de la communication comme les homographes ou les homophones. Si l'on veut mesurer la distance entre des formes parlées, une transcription des phonèmes ne suffira peut être pas, car elle ne fait pas ressortir les ressemblances entre des caractéristiques audibles comme la sonorisation, le point et le mode d'articulation si bien que l'on ne rendra pas compte des ressemblances entre des paires telles que p,b/, /t,d/, /k,g/. Pour ce faire, on ourra utiliser des comparaisons basées sur les différences entre traits distinctifs. Pour illustrer ces trois possibilités, reprenons un des exemples de mots apparentés déjà cité: l'anglais cotton et l'espagnol algodón.

Distance	Orthographique	Phonologique	Allophonique
	c o t t o n	k o t o n	k o t o n
	a l g o d o n	a l g o d o n	a l g o d o n
	1+1+2+0+1+2+0+0	1+1+2+0+2+2+0	10+10+1+0+1+1,5+1,5
	$D = 7/_{13} = 0,54$	$D = 8/_{12} = 0,67$	$D = 24/_{120} = 0,20$

Après avoir pris la distance syllabique et tenu compte de la différence d'accent, l'utilisation des traits distinctifs dans les comparaisons allophoniques fait que les deux mots se ressemblent beaucoup plus que si l'on ne tenait compte que de leur forme orthographique ou de leur forme phonologique. On remarquera que sur la même échelle qui établit une différence maximum de 20 entre les traits distinctifs de deux sons, n'importe lesquels, de l'anglais et de l'espagnol, si on compare le /k/ anglais et le /g/ espagnol, la distance entre leurs traits distinctifs est vingt fois plus petite

que si l'on jugeait d'après une simple représentation graphique où les phonèmes seraient considérés comme étant entièrement différents.

On pourra utiliser ces tables vectorielles de différences entre traits distinctifs non seulement pour mesurer la différence entre les éléments formels des codes mais aussi pour calculer la distance entre des portions équivalentes de discours.

Comme nous l'avons vu, les différences et les ressemblances au niveau des codes n'entraînent pas nécessairement les différences et les ressemblances correspondantes au niveau des messages équivalents. Si l'on veut une vue complète des différences entre deux langues, il faudra mesurer les différences entre les performances c'est à dire observer les codes en action. En d'autres termes, il faudra chercher à mesurer la distance au niveau du discours.

3.3 Limites des mesures entre codes

Les différences entre codes n'entraînent pas automatiquement des différences correspondantes entre les messages qu'ils produisent. La raison en est que la distribution des ressemblances et des différences, même produites par la juxtaposition de deux codes la plus exhaustive possible ne sera pas forcément identique à celle que l'on trouvera dans deux textes équivalents. Examinons par exemple les énoncés suivants et comparons les ressemblances entre le premier groupe de phrases équivalentes à celles du second groupe.

1er groupe	2e groupe
An: Here comes the old man who is blind; give him your hand.	An: A permanent committee of the government is responsible for changes in the Constitution.
All: Hier kommt der alte Mann der blind ist; gib ihm die Hand.	All: Ein ständiger Regierungsausschuss its für die Änderungen des Staatsgrundgesetzes verantwortlich.
Fr: Voici le vieillard qui est aveugle; donne-lui ta main.	Fr: Un comité permanent du gouvernement est responsable des changements dans la Constitution.

Il est évident que dans le 1er cas, le texte anglais est beaucoup plus proche de l'allemand que du français, tandis que dans le second cas, c'est l'inverse qui est vrai: le texte anglais étant plus proche du texte français.

Deux langues peuvent ainsi avoir un certain nombre de ressemblances à un niveau mais pas à un autre. Il peut y avoir des ressemblances dans le vocabulaire mais non dans la grammaire; des ressemblances dans la grammaire mais non dans le vocabulaire; des ressemblances entre certains traits de prononciation mais non pas entre d'autres traits. Par exemple, le français et le hongrois sont sur la plupart des points tout à fait différents; mais ils se ressemblent beaucoup sur certains traits de

prononciation: des phonèmes (des voyelles d'avant arrondies) et des prosodèmes (égalité syllabique). De la même façon, on dit que l'anglais et le chinois sont des langues très différentes mais pourtant sur certains points, l'anglais est plus proche du chinois que de l'allemand: d'après certains sondages effectués par Jespersen, il semblerait que l'ordre des mots (Sujet - Verbe - complément d'objet) dans la prose de l'anglais moderne ait une régularité qui varie en gros entre 90% et 97%, ce qui est voisin du chiffre que l'on trouve pour le chinois (près de 100%) mais très éloigné de l'allemand pour lequel la régularité ne dépasse guère 30% (Jespersen 1912).

A toutes fins pratiques, on observera la ressemblance entre deux langues d'après les messages ou les textes que produiront les langues. Comme les chercheurs qui s'occupent de l'apprentissage des langues et du bilinguisme ont affaire à des cas particuliers il est important de mettre au point des techniques pour traiter ces cas. Pour cela, on pourra analyser et mesurer non seulement les différences entre les codes mais également la distance entre les messages ou les échantillons de discours que produisent les codes dans des contextes comparables.

Les opérations nécessaires pour entraîner cette équivalence dépendront du degré de différence entre les deux langues, c'est à dire la distance entre elles, et donc, on pourra les utiliser pour mesurer cette distance. Ces opérations vont du simple emploi d'homophones et homographes qui sont identiques dans les deux langues à des changements de structure et de niveau très complexe.

Il est donc possible d'établir une échelle qui reflètera cette différence. Si nous posons comme postulat l'équivalence entre les significations, nous mesurons des différences de forme et de fonction. Nous pouvons partir soit de la forme orthographique, soit d'échantillons de discours transcrits phonétiquement. Pour des raisons de simplicité et d'uniformité nous commencerons par prendre les textes orthographiques.

L'équivalence de deux textes orthographiques juxtaposés pose deux problèmes. A l'intérieur de chaque langue, il y a une différence entre ce qui doit être dit et ce qui peut être dit. Dans certaines langues, il peut n'y avoir qu'une seule façon d'exprimer un message donné; dans d'autres langues, il peut y avoir de nombreuses façons d'exprimer le message équivalent. Dans une langue, il peut y avoir plus de liberté que dans l'autre, et cette différence de degré de liberté constitue elle-même une mesure de la distance entre les langues. Ce degré de liberté est considérablement restreint dès que le segment de discours doit fonctionner dans le contexte d'une situation (cf 3.2 ci-dessus). Cependant, pour pouvoir comparer deux segments équivalents de discours, quels qu'ils soient et mesurer la distance qui les sépare, on doit d'abord connaître les traits par lesquels un texte peut différer d'un autre texte.

Quand on mesure la différence entre deux langues dans les textes sémantiquement équivalents, on peut, soit analyser et classer les différences en donnant des valeurs à chaque type de différence et c'est une mesure taxinomique, soit aboutir à une mesure unique des textes en les juxtaposant et c'est une mesure intégrale.

Dans le cas de la mesure taxinomique, il se peut qu'il faille aller au-delà de ce qui est perceptible et que l'on doive relier le texte au code qui l'a produit; dans le deuxième cas, on ne mesure que ce qui est patent et directement observable. Nous commencerons par le premier cas: l'analyse de la distance taxinomique dans le discours.

IV

LA DISTANCE TAXINOMIQUE DANS LE DISCOURS

4.1 Les dimensions de la différence Page 105

4.1.1 Les différences de dimensions: typologie 106

4.1.2 Les différences de dimensions: algèbre 109

4.1.2.1 Notation de la hiérarchie 110

4.1.2.2 Notation de l'ordre 114

4.1.2.3 Notation de la longueur 114

4.1.2.4 Notation de la profondeur 115

4.1.2.5 Notation des éléments 115

4.1.2.6 Notation des relations 116

4.1.2.7 Notation de l'ensemble 116

4.1.3 Analyse différentielle 117

4.1.3.1 Nombre de dimensions 117

4.1.3.2 Composantes des différences 120

4.2 Degrés de différence 121

4.2.1 Conversion de la hiérarchie 123

4.2.2	Conversion de l'ordre	125
4.2.3	Conversion de la longueur	126
4.2.4	Conversion de la profondeur	129
4.2.5	Conversion des éléments	130
4.2.6	Conversion des relations	133
4.2.7	Conversion de l'appartenance à un ensemble	134
4.2.8	Distance pluri-dimensionnelle par conversion	135

4- La distance taxinomique dans le discours

Des segments de discours équivalents peuvent différer de plusieurs façons, du fait que toutes les langues disposent de divers moyens d'expression tels que l'intonation, le vocabulaire et l'ordre des mots. A l'intérieur de chacun de ces domaines, de chacune de ces dimensions, il y a de grandes différences entre les langues, à la fois de type et de degré. Quand on procède à des mesures taxinomiques on doit d'abord définir les différences de type à l'intérieur de ces dimensions de différence entre les langues avant de mesurer les degrés de différence.

Afin de mettre au point des mesures pour les divers types de différence, on devra d'abord définir ces types. Il faut commencer par se poser cette question fondamentale: De quelle façon deux langues peuvent-elles différer dans les textes qu'elles produisent? Il faut déterminer les dimensions de la différence avant de mesurer le degré de différence.

4.1 Les dimensions de la différence

Pour étudier les dimensions de la différence entre langues au niveau du discours, on commencera par isoler les divers types, puis on élaborera une façon commode pour y renvoyer. En d'autres termes, avant de procéder à une forme quelconque de mesure taxinomique, il faut disposer d'une typologie et d'une algèbre des différences linguistiques.

4.1.1 Les différences de dimensions: typologie

De quelle façon les langues peuvent-elles différer par la forme? Elles devraient aussi pouvoir différer par les dimensions communes à toutes les langues. Quelles sont ces dimensions? Il semble évident comme nous l'avons vu que, de par leur nature même, toutes les langues doivent disposer de certains types d'éléments (les sons, les mots, les phrases) dont certains sont plus inclusifs que d'autres, produisant ainsi à une hiérarchie d'inclusion. Dans le discours, ces éléments doivent apparaître dans un certain ordre, l'un précédant l'autre. Ces séquences d'éléments doivent obligatoirement avoir une certaine longueur, certaines étant plus longues que d'autres. Les éléments qui naturellement appartiennent à des classes doivent aussi avoir une certaine profondeur, certains permettant d'effectuer plus de substitutions que d'autres. Ces éléments sont composés d'unités perceptibles visuellement ou accoustiquement, étant identiques ou équivalents. Inévitablement certains de ces éléments auront des relations avec d'autres éléments; et finalement tous les éléments font partie d'ensembles dont les membres ont des fonctions similaires.

Deux textes linguistiques équivalents peuvent différer sur l'une ou plusieurs ou l'ensemble de ces dimensions. Le nombre d'échelons de leurs hiérarchies peut être différent comme c'est le cas quand on compare une langue très agglutinante comme l'eskimo et une langue très analytique comme l'anglais. L'ordre obligatoire ou facultatif de leurs éléments, fonctionnant comme des unités à l'intérieur d'une structure peut

aussi être différent: il suffit de comparer l'ordre des mots en français, en allemand, en anglais et en latin. La longueur, mesurée d'après le nombre d'éléments dans des unités correspondantes, peut aussi être différente; ceci apparaît clairement si on compare des mots en chinois et en hongrois. Cependant, il est même possible que certains éléments n'aient pas d'équivalent dans l'autre langue; d'autres peuvent ne pas apparaître dans le discours si bien que leur longueur est zéro, d'autres peuvent avoir un équivalent auquel s'ajoute quelque chose par exemple à un verbe d'une langue correspond dans l'autre langue un verbe et une locution. Dans certains cas, cet équivalent qui dans la hiérarchie est située à un niveau différent comme lorsque un pronom se rend par un groupe de mots dans une autre langue) entraîne de ce fait même une différence de longueur. Des textes équivalents peuvent aussi différer par la profondeur mesurée d'après le nombre d'unités qui peuvent permuter avec chaque unité de l'énoncé. Les unités pronominales du japonais ou du bengali présentent plus d'alternatives que celles de l'anglais ou du néerlandais. La profondeur est une mesure de ce que l'on peut considérer comme équivalent. De la façon la plus manifeste, la configuration des éléments eux-mêmes, la forme des morphèmes (mots et terminaisons) diffère d'une langue à l'autre et pourtant le contact entre langues et la parenté génétique sont la cause de nombreuses ressemblances et de nombreuses similitudes qui tendent à réduire cette sorte de distance entre langues: des similitudes visuelles (par exemple des homographes comme le mot nation en français et en anglais) et des similitudes auditives (par exemple des homophones comme /haus/ en anglais et en allemand). Des textes équivalents en deux langues peuvent aussi différer en ce qui concerne le nombre de relations

spécifiées entre les unités, certaines langues comme l'allemand réclamant des changements pour faire accorder la forme de l'article et de l'adjectif avec celle du nom; d'autres langues comme l'anglais n'exigeant pas ce type de relation formelle. Finalement deux langues peuvent différer par leur façon d'organiser leurs éléments en ensembles distinctifs tels que les classes formelles, comme lorsque quelque chose qui s'exprime par un verbe en français doit apparaître en anglais sous la forme d'un adjectif. Ainsi l'organisation formelle de deux langues pourra différer sur l'une, plusieurs ou l'ensemble des sept dimensions suivantes: la hiérarchie (H), l'ordre (O), la longueur (L), la profondeur (P), l'équivalence entre éléments (E), les relations (R) et l'appartenance à un ensemble (A).

La première opération taxinomique consiste à déterminer la (ou les) partie(s) de deux segments de discours sémantiquement équivalents qui a (ont) des significations équivalentes. On peut opérer en juxtaposant les textes et en les découpant en unités équivalentes, par exemple

Fr	Il	se lave	le matin
An	He	washes	in the morning

Quand on a ainsi découpé les deux textes équivalents de telle sorte que les unités équivalentes soient comparables, on peut examiner chaque paire pour rechercher les différences sur chacun des sept points ci-dessus.

Pour certaines langues, les unités équivalentes ne différeront que sur une seule dimension, pour d'autres langues sur deux, d'autres trois, d'autres encore sur toutes les sept. Par exemple les pronoms

personnels de l'anglais et de l'allemand bien qu'ayant la même profondeur diffèreront en ce qui concerne les relations obligatoires, tandis qu'en français et en anglais, ils diffèreront à la fois sur la profondeur et les relations obligatoires; dans le premier cas nous avons une différence à une dimension, dans le second, à une différence à deux dimensions. Ainsi les traits de deux langues comparés ou contrastés de cette façon pourront différer suivant de nombreuses possibilités depuis une simple différence de hiérarchie (H) jusqu'à sept différences simultanément (HOLPERA) (Tableau 9).

On trouvera ci-dessous des exemples de différence sur une, deux, trois, quatre, cinq, six et sept dimensions ainsi que des exemples montrant un certain nombre de combinaisons de ces composantes. Afin de pouvoir identifier, au-delà des mots ou groupes de mots particulier, le type d'agencement et de relation en cause, il faut construire un système pour noter les différences essentielles. En effet à l'intérieur de chacune des dimensions les différences ne sont pas toutes du même type et n'ont pas la même portée. Afin de les traiter dans les cas spécifiques, il faut mettre au point une sorte de notation simple mais adéquate, une sorte d'algèbre au sens d'une utilisation systématique et généralisée de symboles pour exprimer et analyser la relation qui existe entre des concepts.

4.1.2 Les différences de dimensions: établissement d'une algèbre

Si on veut élaborer un système simple de notation pour analyser les différences linguistiques pour chacune des dimensions ci-dessus, il faut garder présents à l'esprit les symboles qui existent

déjà en linguistique et en mathématique, tout en essayant de résoudre les difficultés avec un maximum de simplicité et d'élégance. On peut y arriver avec l'aide d'un modèle simple de deux textes différents mais équivalents (figure 2).

Si les deux langues sont symbolisées respectivement par A et X et les unités de leurs textes respectivement par ABC et XYZ, les éléments constitutifs de chacune de ces unités peuvent être /abc/ et (xyz) de façon à ce que l'unité A soit équivalente (=○=) à l'unité X B à Y et C à Z (tableau 10).

Avec cette notation de base, on pourra exprimer les différences de hiérarchie d'ordre, de longueur, de profondeur, d'équivalence ou de similitude des éléments, de relation et d'appartenance à un ensemble de deux textes équivalents. A l'aide de notations supplémentaires, on pourra tenir compte de certaines variables propres à certaines dimensions.

4.1.2.1 Notation de la hiérarchie

Si l'on symbolise la hiérarchie à l'aide de notations souscrites quasi-conventionnelles, il suffit d'un nombre limité de marques pour décrire toute équivalence ou toute équivalence de rang, en partant de la phrase jusqu'à l'allophone. Ces marques seront les suivantes: la phrase ($A_{\circ}$), la proposition (A:), le syntagme (A;), le groupe de mots ($A_{,}$) le mot (A.), le morphème (-A), le prosodème ($\bar{A}$), le phonème (/A/) et l'allophone ([A]). Par exemple, dans les textes équivalents que voici, un segment, le syntagme français (A;) n'importe qui est équivalent

TABLEAU 9

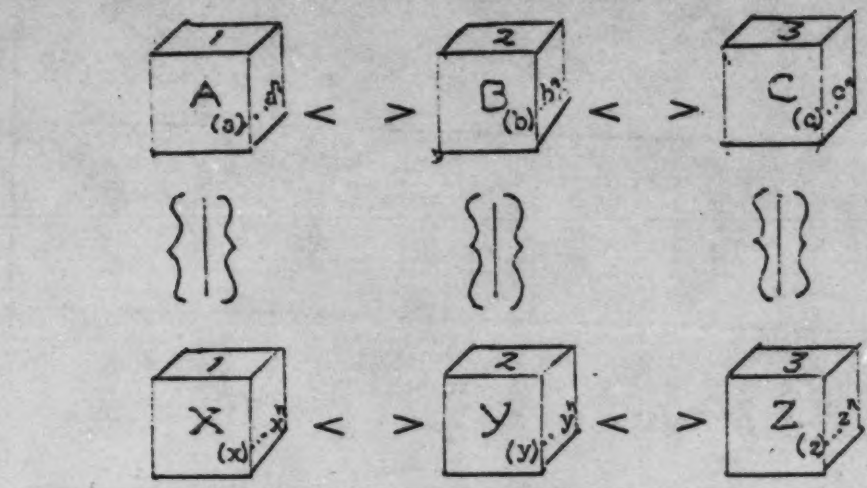
Dimensions de la différence entre langues

1-D	2-D	3-D	4-D	5-D	6-D	7-D
H	HL HO	HLR EAO ERA	HOLP AEOR AOPR	OPERA LPERA	OLPERA	
O	OD OR	HOL LAE LPA	HORL AEPR AOLP	HOPEA OLPRA PERAH		
L	OL PR	HOP PAE HPA	HOPR AELP AHLP	LERAH HOLER HELPA	LPERAH	
P	LR LP	HOR RAE HLE	OLPR AEHP HLRE	PRAHO ERAHO	PERAHO	
E	HR HD	LPR HLA O E	HLPR AHLR HOPE	HLPRA OLERA		HOLPERA
R	HE HA	OPR OPS OLE	AEHL AHOP LPRE	EAHOL RAHOL	ERAHOL	
A	OE OA	OLR OLA LRE	AEOP ALPR OLRE	RHOLP AHOLP	RAHOLP	
	LE LA	OLP LRA HRE	AEOL AOLR RHPE	AOLPE HOLPE		
	PE PA	RHP HRA HOE	AELR ARHP HOLE		AHOLPE	
	RE RA	HLP HOA ORE	AEHR AHOL HORE	HLPER OLPER	HOLPER	
	AE	HEA ORA PRE	AEHO AHOR OPRE	HOPRE		
		LPE HPE	OLPE HLPE			

Types de différences en ce qui concerne: la Hiérarchie (H), l'Ordre (O), la Longueur (L), la Profondeur (P), les Eléments (E), la Relation (R), l'Appartenance (A) à un ensemble.

FIGURE 2

Modèle différentiel



Modèle algébrique de textes équivalents en langue A et X montrant la hiérarchie (Aa), l'ordre ABC), la longueur (1-2-3), la profondeur (a^n), les éléments (a,b,c), la relation <A,B> et l'appartenance à un ensemble {A,X).

Exemple	A	B	C
	une	grande	table
	|	|	|
	a	big	table
	X	Y	Z

Equivalence (≍) :

A ≍ X, B ≍ Y, C ≍ Z.

Congruence (≡)

A ≡ X, B ≡ Y, C ≡ Z

Identité (≡)

C ≡ Z

Différence (~)

(ABC) ~ (XYZ) = A ≠ X ± B ≠ Y

TABLEAU 10

Signes et Symboles pour les mesures taxinomiques

RANGS		UNITES		CATEGORIES			COMBINAISONS
$A_{\circ}$	phrase	A	X	N nominal	n	pronom	(a,a) constituants
A:	proposition	B	Y	V verbal	v	auxiliaire verbal	
A;	syntagme	C	Z	D adjectival	d	déterminant	<a,b> relations
A,	groupe			A adverbial	a		
		ELEMENTS					
A.	mot			c conjonction	p	préposition	{a,b} ensembles
-A	morphème	a	x				
$\bar{\bar{A}}$	prosodème	b	y	RELATIONS			
A/	phonème	c	z	Equivalence	≂		}a,b{ non-ensembles
A]	allophone			Congruence	≅		
				Identité	≡		
				Egalité	=		
				Différence	~		

pour ce qui est du sens à un schéma intonatif anglais ($\bar{X}$) descendant-montant ($\vee$)

Fr Ils n'invitent pas n'importe qui /

An They don't invite $\vee$ anybody A, ⌒ $\bar{X}$

4.1.2.2 Notation de l'ordre

Si A est équivalent en sens à X, B à Y et C à Z, on peut indiquer l'identité ou la différence d'ordre simplement par la suite des lettres. Ainsi le mot français répond (A) est équivalent au mot anglais answers (A) et Fr toujours (B) est équivalent à An always (Y) dans les phrases suivantes:

Fr Il répond toujours /

An He always answers A.B ⌒ Y.X

4.1.2.3 Notation de la longueur

On indiquera simplement la longueur par le nombre d'unités. Toutefois, pour éviter la confusion avec les différences de hiérarchie, il sera bon de rester à l'intérieur des limites du rang considéré, car plus le rang est élevé, plus la séquence des unités est susceptible d'être longue mais ce n'est pas forcé: Merci peut constituer une phrase au même titre que je vous remercie beaucoup. Dans l'exemple qui suit, un verbe français formé d'un seul mot est exprimé par deux mots en anglais:

Fr Chercher /

An Look for A. ⌒ X.Y

On peut aussi indiquer la longueur en numérotant chaque élément constitutif (voir ci-dessous).

4.1.2.4 Notation de la profondeur

On peut considérer que la profondeur est le nombre d'équivalents que le code permet de substituer, non seulement le nombre d'équivalents de vocabulaire différents mais aussi le nombre d'équivalents du niveau du syntagme, de la proposition et de la phrase, bref le nombre de façons différentes dont on dispose pour établir un texte équivalent. Dans l'exemple suivant, on peut rendre le mot français lueur de deux façons en anglais: par glow ou par gleam.

Fr	Lueur /	
An	Glow	A ⊂ X. Y.
	Gleam	

On peut aussi noter la profondeur en portant en indice le nombre d'alternatives. Cette méthode est particulièrement utile pour mesurer les degrés de différence $A. = X^2$.

4.1.2.5 Notation des éléments

Afin de noter les ressemblances et les différences d'éléments constitutifs de deux textes équivalents on pourra utiliser une lettre minuscule pour renvoyer à l'unité ou au segment que l'on est en train d'analyser: (a) pour A, (X) pour X. Par exemple, dans le groupe de mots français (A), pauvres parents, le premier mot pauvres se notera comme le premier constituant (a_1) et le second mot parents comme le second (a_2)

en numérotant de gauche à droite. On comparera aussi en allant de gauche à droite, si bien que si, dans la partie droite de l'équation apparaît un élément qui est identique à son équivalent de la partie gauche, on pourra l'indiquer de cette façon, comme on indique l'identité orthographique du mot parents dans les deux textes

Fr	pauvres	parents /		A.B ⌯ X.B
An	poor	parents	ou	$A(a_1a_2)$ ⌯ $X(X_1a_2)$

4.1.2.6 Notation des Relations

Toute relation qu'on devra indiquer pourra être exprimée entre les signes conventionnels marquant la relation: <> mais on ne notera que les différences significatives. Par exemple:

Fr	grande maison /		
An	big house	<A.B.> ⌯	X.Y

On pourra ainsi indiquer de nombreux types d'accord et de rection avec divers degrés de raffinement.

4.1.2.7 Notation de l'Ensemble

Si deux unités ou deux éléments ne font pas partie du même ensemble ou d'un ensemble équivalent (classes formelles, parties du discours, catégories grammaticale, type de syntagme etc...) on pourra indiquer ce fait à l'aide de la parenthèse employée conventionnellement pour définir un ensemble {}. On pourra affecter le type de catégorie d'un indice pour indiquer qu'elles appartiennent à des ensembles différents (par

exemple N pour nominal, V pour verbal) $\{A_\circ\}_N \rightleftharpoons \{X_\circ\}_V$ ou bien mettre les parenthèses à l'envers par exemple Fr. attention / An. look out $\}A.\{ \rightleftharpoons \}X.Y.\{$

4.1.3 Analyse différentielle

Une paire equivalente peut différer seulement sur une dimension, sur deux dimensions, sur trois ou plus. La complexité de la différence par conversion sera refletée par la complexité de la formule.

La notation proposée ci-dessus suffit à indiquer les différences dans toute combinaison de dimensions; comme l'ont montré les exemples précédents, on peut le faire à la fois pour la description du nombre de dimensions différentes et pour l'inclusion des diverses composantes dimensionnelles.

4.1.3.1 Nombre de dimensions

Différences à une dimension

Fr Oui, effectivement je suis adulte /

An Yes, I' am an adult

A effectivement / ' $\bar{X}$

H (A/X) : A. $\rightleftharpoons \bar{X}$

Différences à deux dimensions

Fr Ce fut le plus grand que l'on ait jamais vu /

An It was the largest ever seen

A ($(a_1a_2a_3a_4a_5)$): que l'on ait jamais / ever X(x).

B (b) vu / seen Y(y)

He (A/X): A($a_1a_2a_3a_4a_5$): B(b). ⊂⊃ X(x). Y(y).

Différences à trois dimensions

A (a). All Bitte (avec geste) / An. Have a seat $X(x_1x_2x_3)$.

B (b). Danke / Thanks Y(y).

C (c). Bitte / Nil

HEL (A/X) : A (a). B (b). C(c). ⊂⊃ $X(x_1x_2x_3)$. Y(y).

Différences à quatre dimensions

Fr Son frère a le nez long

An His brother has a long, protruding nose

A $\{a.\}_{d_1}$	le	/	a	X $\{x\}_{d_2}$
B <b a>	nez	/	nose	Y (y)
C (c).	long	/	long, protruding	Z (c) Z (z)

POAE

A $\{a.\}_{d_1}$ B<b a> C (c) ⊂⊃ $X\{x\}_{d_2}$ Z (c). Z (z). Y(y)

$\{A.\}_1$ B <a>. C ⊂⊃ $\{X.\}_2$ Z (c). Z (z) Y.

Différences à cinq dimensions

Fr Le chien était en train d'enlever avec la langue du chocolat qui se trouvait sur sa main

An The dog was licking chocolate off his hand

	était en train de /	was .... ing	X(x.-y)
B(b;)(c,){b:}$_v$	d'enlever avec la langue /	lick ... off	Y(y.)z.{y}$_a$
C,	qui se trouvait sur /		
	du chocolat	chocolate	Z.

HOLERA (A/X) A; B(b;) c, {b:}$_v$ ⌯ X(x.-y) Y(y.)z.{y:}$_a$

Différences à six dimensions

Fr On dit que le roi Abdul est depuis longtemps parti /

An King Abdul has allegedly left long ago.

A:	On dit que /	alledgedly	X.
B ($b_1 b_2 b_3$),	le roi Abdul /	King Abdul	Y($y\ b_3$),
C (c_1) $\{c\}_{b_2}$ $\langle c_3 b_1 \rangle$;	est depuis longtemps parti	left long ago	Z(z) $\{z,\}_b$;

or

C(c_1). $\{c_2 z_2\}$. $\langle c_3, c_1\ b_1 \rangle$.

or

C c_1 c_2 c_3 $\langle c_1\ b_1 \rangle$. Z z_1. $\{z_2,\ c_2,\}$.

HOLERA (A/X): A:B ($b_1 b_2 b_3$), C c_1 c_2 $c_3 \langle c_1 b_1 \rangle$; ⌯ X. Y ($y\ b_3$), Z $z_1 \{z_2 c_2\}$.

Différences à sept dimensions

Fr **Le professeur Martin aurait été congédié sans cérémonie**

An **It seems that Professor Martin has been summarily fired (let off).**

A($a_1 a_2 a_3$)	Le professeur Martin /	Professor Martin	X ($x\ a_3$)
-3	-ait . /	it seems that .	Y:

$C(c_1\text{-}b)$: $c_2 \langle c_3 a_1 \rangle \{C_3,\}_N$ aur- été congédié / has been summarily fired (let off) sans cérémonie $Z\left(z, \{z\}_n z^2\right)$

MOLPERA (A/X): $A\ (a_1 a_2 a_3)$ $C\ (c_1\text{-}b)$: $C_2 \langle c_3 a_1 \rangle \{C_3,\}_N \Leftrightarrow Y;\ X\ (x a_3),\ Z\left(z, \{z\}_n z^2\right)$

4.1.3.2 Composantes des différences

Dans chacun des groupes dont nous avons donné un exemple (de une à sept dimensions) les composantes réelles peuvent être fort différentes: ainsi une différence à deux dimensions peut être une différence de hiérarchie et d'ordre (HO), de longueur et relations (LR) ou une des 10 autres combinaisons possibles de deux éléments à partir des sept dimensions

L $\bar{A} \Leftrightarrow \bar{A}\,\bar{B}$ F. Mayonnaise/ J.Mayonezu F. (s) = J. (ss)

O $A.B. \Leftrightarrow B.A.$ F. crédit social / E. social credit

AE $\{A.\}_{d_1} \Leftrightarrow \{X.\}_{d_2}$ F. Levez la main./E. Lift up your hand.

HE $A; \Leftrightarrow X.$ F. Un vrai sauvage/E. He's just a barbarian.

AHE $\{A,\}_N\ B.C; \Leftrightarrow \{X,\}_n Y,Z.$ F. Le temps est a la pluie. / E. It looks like rain.

LEO $A,B,A, \Leftrightarrow X.Y.$ F. Il se rend chez-lui en voiture./E. He drives home.

LEA $A,B; \Leftrightarrow X.$ F. Je lui ai donné de grands coups dans le dos/ / E. I thumped him on the back

PEA $\{A.\}_o \{B.\} \Leftrightarrow \{X.\}_n$ F. Il se rendit fier,orgueilleux./E. He surrendered proudly.

REL $A,B.C: \Leftrightarrow X,Y.Z \langle x.z \rangle$: F. Il n'était pas prêt,n'est-ce pas? / E. He wasn't ready, was he?

LEOA $\{A.\}_v \{B;\}_N \Leftrightarrow \{Y.\}_v \{X.\}_r$ F. Il monta d'un pas lourd l'escalier./

or:

$A.\{B;\}_N \Leftrightarrow Y.\{X\}_r$ E. He plodded up the stairs.

4.2 Degrés de différence

La différence entre langues n'est pas seulement une question de type mais aussi une question de degré. Par exemple, les différences de longueur varient en fonction du nombre d'unités par lequel un segment de discours dans une langue dépasse le segment correspondant dans l'autre langue.

Dans les questions de mesure, les nombres sont nécessaires, tout comme les symboles. On peut incorporer les nombres en tant qu'indice des éléments constitutifs - $(a_1 a_2 a_3 a_4 a_n)$. On peut utiliser les nombres positifs et négatifs en faisant appel aux opérations simples de l'algèbre ordinaire, chaque terme de l'équation représentant un segment de discours dans l'une des langues: (A = B) et (A - A = 0)

On peut ainsi mesurer la distance par conversion ou la distance différentielle entre deux segments de discours équivalents. Quand on mesure la distance par conversion, il est essentiel de commencer par déterminer avec précision ce que l'on convertit et ce à quoi on aboutit car l'opération dans un sens peut être plus complexe que dans l'autre. On pourrait penser que le point de départ (la langue source) (S) va se placer à gauche dans l'équation et que le point d'arrivée ou langue cible (C) va être à droite mais il n'en est rien, car la différence doit représenter le nombre de distinctions supplémentaires que doit opérer le locuteur de la langue source pour arriver aux distinctions de la langue cible. La formule pour mesurer la distance taxinomique par conversion entre discours doit par conséquent être: $D(S \nearrow C) = \{C\} - \{S\}$ Pour simplifier les calculs tout en accroissant le nombre de séquences possibles, un ensemble

de lettres suffira, en commençant à la première lettre de l'alphabet.

{c} - {s} = C (A.) - S (A.) = A. - A. = 0

Ainsi avec une légère modification du système de notation, on peut mesurer les degrés de différence entre deux textes équivalents pour chacune des sept dimensions. On devra d'abord traiter les textes pour déterminer les segments qui sont équivalents. Voici un exemple de la façon de procéder.

1- On détermine la direction de la conversion, c'est à dire laquelle des langues est la langue source (S) et laquelle est la langue cible (C). On écrit le texte cible avant (à gauche ou au dessus) le texte source (à droite ou en dessous).

2- On découpe le texte en segments équivalents sémantique ment

Fr On dit / qu' il / est parti /. = 3 segments

An He / has / allegedly / left = 4 segments

Nombre total de segments 7

3- On affecte de lettres, de gauche à droite et en suivant les segments de la langue cible (ABC...)

A B C

On dit / qu'il / est parti

4- On donne les mêmes lettres aux segments correspondants de la langue source.

He / has / allegedly / left

B C A C

5- On établit l'équation de base ABC = BCAC

6- On développe l'équation de conversion de la façon indiquée ci-dessous.

7- On exprime les résultats sous forme de puissance à chaque dimension de la différence.

$$H^n \quad O^n \quad L^n \quad P^n \quad E^n \quad R^n \quad A^n$$

De cette façon, on peut calculer séparement la distance par conversion pour la différence de hiérarchie, celle d'ordre, de longueur, de profondeur, d'équivalence des éléments, de relation et d'appartenance à un ensemble; ce qui donne en formule:

$$D\ (S/C) = H\ (C\text{-}S)\ O\ (C - S)\ L\ (C - S)\ P\ (C - S)\ E\ (C - S)\ R\ (C - S)\ A\ (C - S)$$

4.2.1 Conversion de la hiérarchie H (C - S)

Si un élément d'une phrase doit changer de rang dans la hiérarchie pour devenir équivalent à un élément d'une autre langue, on peut utiliser le nombre de rangs ou d'échelons qu'il lui faut parcourir sur l'échelle de la hiérarchie vers le haut ou vers le bas pour mesurer la distance de conversion de hiérarchie. Il faudra que les valeurs numériques attribuées à chaque rang ou à chaque échelon de l'échelle aient quelque relation avec sa puissance d'inclusion. C'est pourquoi on attribuera la valeur la plus grande au rang le plus élevé: le texte ou pour plus de commodité la phrase ($A_\bullet$) et la valeur la plus petite au rang qui a la puissance d'inclusion la plus faible: l'allophone ou la combinai-

son d'allophones (A]). Si nous opérons avec neuf rangs, la phrase ($A_.$) occupera le neuvième rang, la proposition (A:) le huitième, le syntagme (A;) le septième, le groupe de mots ($A_,$) le sixième, le mot (A.) le cinquième, le morphène (-A) le quatrième, le prosodème ($\overline{A}$) le troisième, le phonème (/A/) le second et l'allophone le premier ([A]), tout au bas de l'échelle. La distance par conversion de hiérarchie sera de cette façon représentée par le nombre d'échelons qu'il faut gravir (valeur positive) ou descendre (valeur négative) pour transformer un segment d'énoncé d'une langue en un segment équivalent de l'autre langue.

Par exemple, pour que le schéma intonatif de anytime dans la langue source (S) soit équivalent à un segment de discours de la langue cible (C), c'est à dire à toute heure, nous calculerons le nombre d'échelons qui sépare la cible de la source c'est à dire C - S. Ainsi.

LANGUE D'ARRIVEE (C) - LANGUE DE DEPART (S)

	9 A	A 9
	8 A:	A:8
Je ne suis pas libre à toute heure	7 A;	A;7
	6 A,	A,6
	5 A.	A.5
	4-A	-A 4
	3 $\overline{A}$	$\overline{A}$ 3
	2/A/	/A/2
	1 A]	A]1

I'm not free anytime. = A; -A

H (C - S) = A; - $\overline{A}$ = 7 A - 3 A = 4.

En faisant précéder chaque segment de discours que l'on mesure de la valeur correspondant à son rang, on peut tout simplement soustraire les deux valeurs, comme dans l'exemple ci-dessus.

4.2.2 Conversion de l'Ordre O (C - S)

Si l'on doit modifier l'ordre d'un segment de discours d'une langue pour le rendre conforme avec son équivalent dans l'autre langue, on peut utiliser l'importance plus ou moins grande de la modification pour mesurer la distance de conversion à parcourir. On peut mesurer l'importance de la conversion au nombre de changements de position que l'on doit effectuer et à la différence de position que les équivalents occupent dans chaque séquence. On peut quantifier ces différences simplement en calculant le nombre et l'importance des changements de position nécessaires pour que les équivalents soient analogues. On numérote, dans chaque séquence, les éléments constitutifs après en avoir juxtaposé les équivalents ainsi Fr observation unique / An. unique observation une fois comparés donneraient $A_1B_2 - B_2A_1$. La différence doit bien sûr apparaître en nombres positifs, car l'ordre n'entraîne pas de questions de plus ou de moins, étant exclusivement relatif. Si on fait des mesures par conversion au lieu de mesures différentielles, on peut diviser le résultat par deux ce qui donnerait

$$\frac{\begin{matrix} A_1 & B_2 \\ A_2 & B_1 \end{matrix}}{\frac{1 + 1}{2}} = 1$$

La différence peut être causée seulement par le déplacement d'un seul élément comme dans l'exemple ci-dessus ou l'adjectif se place avant le nom dans une langue et après le nom dans l'autre; ou bien la distance entre

les éléments peut affecter la plupart des mots de la phrase. Exemples:

All. Er hatte die Festung allein gegen einen Angriff verteidigt/
An. He had defended the fortress alone against an attack

Si l'on découpe un texte en fonction des équivalents de l'autre texte et que l'on fait la soustraction entre les numéros d'ordre des éléments, on obtient le résultat suivant:

	1	2	3	4	5	6	7	8	9
A: All	Er	hatte	die	Festung	allein	gegen	einen	Angriff	verteidigt
X: An	He	had	the	fortress	alone	against	an	attack	defended
	1	2	4	5	6	7	8	9	3
	0	0	1	1	1	1	1	1	6

= 12

	1	2	3	4	5	6
A: An	`	Any	attempt at	concealment from	her	will
X: Fr	Ce	sera	évidemment	une perte de	temps	que de chercher à
	1	6	7	9	8	3
	0	4	4	5	3	3

	7	6	8	9
	obviously	be	time	wasted
	lui		cacher	quoi que ce soit
	5		4	2
	2		4	7

= 32

(On peut également exprimer les résultats en pourcentage, ou comme une proportion,
ex. 0(an - all)=12/18=2/3=66
ou 0(fr - an)=32/18=7/3=2.

4.2.3 Conversion de la longueur L (C - 8)

Quiconque a comparé une traduction avec l'original est frappé par le fait que l'une des versions occupe plus de place que l'autre ou encore que parfois un seul mot de la langue source a été rendu par plusieurs mots dans la langue cible. On mesure les différences de longueur en termes de constituants immédiats: la longueur des mots en syllabes, en phonèmes ou en graphèmes - A (a b c)., la longueur des groupes de mots en mots A (a.b.c), et ainsi de suite. Tel n'est pas le cas pour les différences de hiérarchie dont certaines entraînent aussi des différences de longueur. On peut mesurer la différence pour chacun des segments équivalents en comptant les unités constitutives immédiates (phonèmes ou lettres, syllabes, morphèmes, groupes de mots, syntagmes, propositions ou phrases) et en faisant une soustraction entre les deux nombres.

Il existe une façon simple de calculer la différence de longueur: c'est d'effectuer une soustraction entre les valeurs des indices finals de chaque segment équivalent. Par exemple pour transformer (F. - An.) le mot français le vieillard en anglais (the old man) on a besoin d'un mot supplémentaire. L (An. - F.) = A

L (An. - F.) = A $(a_1a_2a_3)$, - A (a_1a_2), = $a_3 - a_2 = 3 - 2 = 1$

Si l'on opère la conversion dans l'autre sens, on aura

L (F. - An.) = A (a_1a_2), - A $(a_1a_2a_3)$, = $2 - 3 = 1$ soit un mot de moins.

Les différences de longueur peuvent être contigues (___) ou non contigües (_____), c'est pourquoi il est nécessaire de juxtaposer les deux textes pour obtenir l'équivalence optimale afin de pouvoir mesurer avec précision toute différence de longueur (voir ci-dessous). Comparons par exemple l'anglais post à un des équivalents en français (les postes), et en japonais (posutu)

Fr	post|es = 6		J	pos|u|t|u = 6	
An	post|	= $\frac{4}{2}$	An	pos| |t|	= $\frac{4}{2}$
L (Fr - An).	= 2		L (J - An).	= 2	

De cette façon, les différences d'ordre qui ont déjà été mesurées par une autre formule ne se confondront pas avec les différences de longueur.

Toutefois, quand on analyse deux segments, il y a toujours la possibilité d'une confusion entre la longueur et la hiérarchie. Exa-

minons le second mot de la phrase hongroise suivante: "En hotelban vagyok, nem rokoraimnal" (Je suis à l'hôtel, non pas avec mes parents); même si dans ce type de phrase on utilisait un autre mot tel que "szallodaban", cela ne ferait toujours qu'un seul mot, mais cependant, cela ferait deux mots pour l'équivalent en allemand (im Hotel) et trois mots pour le français (à l'hôtel). Comparons

H	hotelban			= 1 (An. - H.) = 3 - 1 = 2
All	im		Hotel	= 2 (An. - All.)= 3 - 2 = 1
Fr	à	l'	hôtel	= 3 (An. - Fr.) = 3 - 3 = 0
An	in	a	hotel	= 3

C'est pourquoi on pourrait croire que:

L (An. - H.); = 2 L (An. - All.); = 1 L (An. - Fr.); = 0

Toutefois, il y a quelque chose de fondamentalement différent entre d'une part la structure du hongrois et de l'autre, celle du français, de l'anglais et de l'allemand. La différence entre les équivalents est manifestement d'un autre ordre: la dimension de hiérarchie. Car l'équivalent hongrois est exprimé dans le cadre de la structure des mots. On pourra soit se satisfaire de cette distinction et exprimer la différence sous la forme suivante H (An. - H.) = 7 A - 5 A = 2, soit marquer la différence de longueur L (An. - H.) = A ($a_1a_2a_3$) - A (a_1) = a_3a_1 = 2

Comme les indices des constituants renumérotés à la suite ($a_1a_2a_3$) pour chaque langue ne permettent pas de distinguer entre les constituants d'une langue et ceux de l'autre, on ne peut pas maintenant utiliser le même numéro pour représenter des constituants différents. Si l'on veut utiliser les mêmes indices à la fois pour mesurer la longueur et l'équivalence des

éléments, il faudra que la séquence des éléments constitutifs de la langue source suive la séquence de ceux de la langue cible A ($a_1a_2a_3$) - A (a_4a_5). Par conséquent, on juxtaposera les séquences de gauche à droite en affectant chaque constituant de la valeur un (voir ci-dessous 4.2.5).

4.2.4 Conversion de la profondeur: P (C - S)

Quiconque a essayé de traduire un texte dans une autre langue a certainement été embarrassé par le nombre de façons d'exprimer la même chose dans la langue d'arrivée, ou langue cible. Certaines langues sont plus riches que d'autres pour certains types d'expression; elles pénètrent pour ainsi dire plus profondément dans le sujet.

Il est difficile de mesurer les différences de profondeur entre deux segments de discours équivalents, car les alternatives sont rarement exprimées. Dans de nombreux cas, il faut dépasser le texte et remonter au code pour mesurer les possibilités. Plus les possibilités sont nombreuses, plus la langue est puissante dans le domaine d'expression considéré. En effet, on peut considérer ces possibilités comme des puissances que l'on peut additionner ou soustraire: on peut soustraire le nombre total de puissances de la langue de départ de la valeur, exprimée en puissances, du segment correspondant de la langue d'arrivée. Ceci donne une mesure de la distance entre profondeurs. Examinons l'exemple suivant:

Fr:	Je l'ai vu luire	/ An:	I saw it	shine glimmer gleam glitter shimmer glint glisten

Soit P (An. - Fr.) = $A.^7 - A.^1 = 6$

On peut appliquer la mesure de la profondeur non seulement aux différences entre mots mais aussi aux différences à n'importe quel niveau de la hiérarchie.

Comme la distance par conversion ne mesure que dans une seule direction à la fois, il ne faudra pas conclure du fait que la langue cible est capable d'exprimer la même chose de plusieurs façons que la langue source ne dispose que d'une façon de l'exprimer. Pour trouver la différence entre les potentiels de chaque langue, il faudrait faire la soustraction entre le nombre de façons possibles d'exprimer la même chose dans une langue et le nombre de façons de l'exprimer dans l'autre langue. (Voir 3.1 et 3.2)

4.2.5 Conversion des éléments E (C - S)

Les personnes qui apprennent une langue étrangère sont parfois agréablement surpris de constater que certains mots ont la même forme et le même sens dans les deux langues. Ainsi celui qui apprend le français ou l'anglais trouvera que des mots écrits comme page, pipe, nation sont familiers. La présence de mots de ce genre qui sont, en contexte, identiques pour ce qui est de la forme et du sens, facilitera le passage d'une langue à l'autre en ce qui concerne les habitudes de compréhension, car la distance par conversion s'en trouvera réduite. Comment mesurer cette distance par conversion? La façon la plus simple serait de faire la soustraction entre le nombre de constituants identiques et le nombre de constituants différents en prenant soin de ne considérer que les constituants immédiats des segments comparés. Par exemple:

Fr. à son hôtel / An. at his hotel =

$A (a_1\ a_2\ a_3)$, = $A (a_4\ a_5\ a_3)$ =

$a_1\ a_2\ a_3$

$a_3\ a_4\ a_5$

$1 + 1 + 0 + 1 + 1 = 4$ = E (Fr. - An.) = 4

Toutefois, la valeur d'un mot identique dans deux segments équivalents dépend de la longueur des segments: un sur trois fait que les textes se ressemblent plus que un sur six. Comme les segments peuvent ne pas être égaux en longueur, on pourra exprimer le nombre d'éléments juxtaposés identiques en fonction du total des éléments des deux textes. Exemple

Fr. à la page six / An. on page six

$A (a_1\ a_2\ a_3\ a_4)$ = $A (a_5\ a_3\ a_4)$

$a_1\ a_2\ a_3\ a_4$ = 4

$a_3\ a_4\ a_5$ = 3

$1 + 1 + 0 + 0 + 1 = 3$ 7 E (Fr. - An.) = 3/7 = 0,43.

Comme nous calculons en termes de constituants immédiats, nous ne tenons compte que des similitudes au niveau de ces constituants c'est à dire qu'un mot est soit identique soit différent. En dernière analyse, pour tenir compte de toutes les ressemblances, il faut traiter les constituants. Exemple:

Fr une longue planche / An a long plank

$$A_1B_1C_1 = A_2B_2C_2$$

$$A_1 - A_2 = A_1 (a_1a_2a_3) - A_2 (a_4) = 4/4 = 1$$

$$B_1 - B_2 = B_1 (b_1b_2b_3b_4b_5b_6) - B_2 (b_1b_2b_3b_4) = 2/10 = 0,20$$

$$C_1 - C_2 = C_1 (c_1c_2c_3c_4c_5c_6c_7) - C_2 (c_1c_2c_3c_4) = 4/12 = 0,33$$

Si bien que $E(\text{Fr.} - \text{An.}) = \text{Fr.}(A_1B_1C_1) - \text{An.}(A_2B_2C_2) = \frac{1 + 0,20 + 0,33}{3} = \frac{1.53}{3}$

$= 0,51$

4.2.6 Conversion des Relations: R (C - S)

Deux segments de discours équivalents peuvent être identiques en ce qui concerne la hiérarchie d'ordre, la longueur et la profondeur de leurs éléments constituants mais pourtant différer profondément sur le nombre des relations obligatoires entre les éléments. On peut mesurer ces contraintes en comptant le nombre pour chaque segment et en opérant une soustraction entre les deux totaux.

_	le vieux cheval		the old horse
Français	la vieille maison	/ Anglais	the old house

Fr (2) - An (0) = 2.

En d'autres termes, pour chacun des deux segments de discours équivalents de trois mots, le français a deux contraintes de plus que l'anglais. Soit R (Fr - An) = 2 . En utilisant la même notation que ci-dessus, on peut indiquer ainsi: < > les relations, chacune s'appliquant à une contrainte du système (voir 4.1.12) On obtiendra le degré de différence en comptant le nombre de ces relations dans chaque terme de l'équation et en faisant la différence. Exemple:

Fr ces vieilles chaussures / An These old shoes

$A\langle a_1 a_2\rangle \langle a_2 a_3\rangle - A\langle a_4 (a_5) a_6\rangle = A\langle 2\rangle - A\langle 1\rangle = 1$

Comme pour la longueur, les relations peuvent être contigues ou non-contigües; on pourra par conséquent les calculer de la même façon.

Fr Où sont ces belles pêches que je vous ai apportées?/
An Where are those nice peaches I brought you?

$A\langle a_1 a_2 a_3\rangle \langle a_2 a_3\rangle, + B(b_1 b_2 b_3 b_4, 2\langle a_3\rangle); - A(a_5 a_6 a_7 \langle a_5\rangle), - B(b_1 b_2 b_3); = 4$

Pour obtenir des mesures plus précises, on pourra classer les relations en catégories, par exemple le nombre (1), le genre (3), le cas (3) etc. Si bien que si on le souhaite, on pourra tenir compte de la différence de type de relation. Exemple:

Fr Ils ont levé <u>la main</u> / An They lifted up <u>their hands</u>

$$\langle A \rangle_2 - \langle A \rangle_1 = 1$$

mais comme Fr <u>la</u> et An <u>their</u> n'appartiennent pas au même sous-ensemble de déterminants, il faudra également tenir compte de ce fait (voir ci-dessous § 4.2.7)

4.2.7 <u>Conversion de l'appartenance à un ensemble</u>: A (C - S)

Si un élément fait partie d'un ensemble ou d'une catégorie (par exemple les nominaux) dans une langue, l'équivalent dans l'autre langue pourra faire partie d'une autre catégorie. Parfois l'équivalent d'un verbe peut être un nom (exemple An <u>Rent a car</u> / Fr <u>Location d'autos</u>, ou l'inverse (An: <u>Adjournment: 5 p.m.</u> / Fr: <u>La séance a été levée à 17 h</u>). En fait, tout élément d'une langue traduit dans une autre langue peut théoriquement faire partie d'une catégorie différente ou d'un ensemble grammatical différent.

Pour mesurer cette dimension, on pourra soit seulement tenir compte du fait que l'ensemble est identique ou non identique soit essayer de mesurer le degré de différence. Dans le premier cas on placera les constituants qui n'appartiennent pas au même ensemble entre parenthèses inversées (non ensemble). Exemple:

An <u>Gargle</u> twice a day / Fr <u>Gargarisme</u> deux fois par jour

$$A. - \}A.\{ = \}1\{$$

Dans le second cas, on pourra mesurer la différence d'appartenance à un ensemble sur toute échelle appropriée portant les différences entre les divers ensembles que chaque langue possède. Exemple

An He was ski-ing / Fr Il faisait du ski

$$\{A\text{-}\}_4 - \{A.\}_1 = 3$$

Pour l'anglais et le français, on peut faire des échelles très précises qui montreront les différences entre les sous-classes de déterminants ou se contenter de distinctions moins fines telles que noms (N) adjectifs (D), adverbes (A), verbes (V) et autres mots structuraux (S). Comparons:

	Anglais		Français	
	N 1	┌──	1 N	J'ai faim
I'm hungry	D 2	──┘	2 D	
	A 3		3 A	
	V 4		4 V	
	S 5		5 S	

$$\{A.\}_2 - \{A.\}_1 = 1$$

4.2.8 Distance pluri-dimensionnelle par conversion

On peut maintenant donner une valeur à chaque dimension ce qui permet d'obtenir un tableau plus exact de l'importance et du degré de la différence qui existe entre deux textes équivalents. Voici plusieurs exemples qui montreront comment on arrive à ce résultat:

L Fr. manicure / J. manikyua s/ss

$\bar{A} = \bar{\bar{A}}\bar{A} = L(1\text{-}2) = L^{-1}$

LE Fr. Ses machoires étaient comme soudées./

En. His jaws were locked.

$A(a_1a_2) - A(a_4) = L(2\text{-}1), E(3) = L^{+1}E^{3}$

Fr. trottiner / E. toddle along

$A(a_1) - A(a_2a_3) = L(1\text{-}2), E(3) = L^{-1}E^{3}$

AHE Fr. Elle resonne lugubrement. /

En. It has a sinister sound.

$\{A.\}_{A} - \{A,\}_{M} = A(3\text{-}1), H(5\text{-}6), E(1+2) = A^{-2}H^{-1}E^{3}$

PEA Fr. Il se tortura à calculer son impôt. /

En. He wracked his brains struggling, straining to figure out his taxes.

$\{A.\}^{1}_{S} - \{A.\}^{2}_{VA} = P(1\text{-}2), E(1+2), S(5\text{-}5) = P^{-1}E^{3}A^{-2}$

LEA Fr. Il se sentit vide de tout. /

En. All he felt was an emptiness.

$A\{a_1a_2a_3\}_{A} - A\{a_1\}_{M} = L(+2), E(4), A(+1) = L^{+2}E^{4}A^{+1}$

LEO E. I like last year's dresses better./

F. J'aime mieux les robes de l'an passé.

$A(a_1).\ B(b_1b_2), C(c_1).\ A(a_1). - A(a_3a_4), C(c_2c_3), B(b_3b_4b_5b_6);$ =

$L(2\text{-}2) + L(2\text{-}4) + L(1\text{-}2),$

$E(12),\quad O(4) = L^{-3}E^{12}O^{4}$

LEOA F. Elle cria au meutre comme une perdue. /

E. She shrieked blue murder.

$A(a_1).B\ (\{b_1\}_s b_2),\ A(a_2a_3a_4);$ -- $A(a_5).B\ (\{b_3\}_o\ b_5),$ =

$L(4\text{-}1),\ E(9),\ O\binom{123}{121} = 2),\ S(5\text{-}2)$ = $L^{+3}E^{9}O^{2}A^{+3}$

L'avantage de ces mesures taxinomiques est qu'elles indiquent non seulement le degré de différence entre les langues tel qu'il se manifeste dans la production du discours mais aussi les dimensions sur lesquelles elles diffèrent. Les mesures indiquent à la fois le combien et le comment. Elles ont cependant un inconvénient: on est obligé de calculer autant de différences qu'il y a de dimensions, chaque mesure, dépendant, d'une classification et d'une interprétation correctes et précises des catégories des langues comparées. On doit connaître la façon dont les codes ont produit les messages correspondants, les autres alternatives possibles, on doit savoir si les relations sont obligatoires ou facultatives, connaître les classes grammaticales qui sont liées et beaucoup d'autres points semblables. En d'autres termes les mesures mesurent plus de choses qu'il n'y paraît.

On pourrait obtenir un type de mesure direct, objectif et plus utile dans certains cas en limitant les différences à ce qui est immédiatement observale. Pour répondre à certains besoins, il serait également bon d'arriver à un seul chiffre au lieu d'avoir sept mesures de différences entre deux textes équivalents. On atteindra ces objectifs en mettant au point des mesures intégrales.

V

LA DISTANCE INTEGRALE DANS LE DISCOURS

5.0	Traitement des textes	Page	140
	5.0.1 Textes orthographiques		142
	5.0.2 Textes phonétiques		146
	5.0.2.1 Segmentation en syllabes		147
	5.0.2.2 Segmentation en phonèmes		151
5.1	Mesures statiques		152
	5.1.1 Segments supra-prosodiques		153
	5.1.2 Segments infra-prosodiques		159
	5.1.2.1 Unités syllabiques		159
	5.1.2.2 Unité phonémique		163
5.2	Mesures dynamiques		166
	5.2.1 Séquences supra-prosodiques		167
	5.2.2 Séquences infra-prosodiques		171
	5.2.3 Intégration		176

5- La distance intégrale dans le discours

Par opposition aux mesures taxinomiques de chaque dimension de la différence qui existe entre deux segments de discours équivalents, on trouve la mesure des différences formelles qui existent entre deux textes suivis. Pour ce second type de mesure, on fait correspondre un texte suivi dans les deux langues de façon à faire ressortir la divergence entre les deux textes sous forme d'une seule mesure.

5.0 Traitement des textes

Cependant, avant d'entreprendre toute mesure systématique, il est utile de traiter les textes de manière à éviter les erreurs d'analyse et de calcul. Le traitement comprend trois opérations principales: la juxtaposition, l'appariement des segments et la numérotation.

Pour commencer, on juxtapose l'un en dessous de l'autre les deux textes puis on les examine pour déterminer exactement les segments qui se correspondent et on le découpe sur cette base. Le découpage en couches successives de constituants immédiats qui se fait couramment pour l'analyse descriptive unilingue n'est peut-être pas le mieux adapté car pour les segments de discours longs ou complexes, il risque de prolonger la phase analytique de l'opération sans apporter grand'chose au produit final. Nous préconisons ici un type de découpage par équivalence

minimale. Cette équivalence minimale représente dans chaque texte les segments les plus petits que l'on peut, sans tenir compte de leur ordre, considérer comme équivalents.

On apparie ensuite par des traits les segments de chaque texte qui se correspondent et on modifie l'espacement si cela s'avère nécessaire. On numérote ensuite les segments appariés pour facilité l'identification. Exemple

Fr	Le patron		a subi		un		dur		coup /
An	The boss		had		a stroke of		bad		luck
		1		2		3		4	5

Ceci aboutit à une juxtaposition de deux séquences de segments ayant une équivalence sémantique minimale. On peut répéter ce processus de découpage jusqu'à ce qu'on ait atteint la plus petite unité comparable. Le point où on s'arrête dépend de la finesse de l'analyse qu'on veut faire. Si c'est le graphème ou la lettre qui constitue l'unité finale, on aura vraisemblablement des résultats sensiblement différents que si on avait choisi le phonème comme la plus petite unité formelle; mais si on compare les langues pour leur apparence extérieure, ce sera peut-être la forme la plus appropriée. Dans ce cas, on comparera des segments de discours dans leur transcription orthographique; dans l'autre cas, il nous faudra des textes dans une transcription prosodique, phonologique ou phonétique. Nous commencerons par étudier le découpage des textes orthographiques avant d'aborder l'étude des textes dans leur transcription phonétique.

5.0.1 Textes orthographiques

Au cours de l'exposé du traitement des textes, nous présenterons les opérations à gauche et les exemples correspondants à droite.

Opérations	Examples
1. Juxtaposer deux textes équivalents A et X	A. Le patron a subi un dur coup. X. The boss had a stroke of bad luck.
2. Marquer les segments équivalents en soulignant (X) ou avec un trait au-dessus ($\overline{a}$) en indiquant l'équivalence d'un trait vertical $\frac{\overline{a}}{\underline{X}}$	1 2 3 4 5 6 A Le patron a subi un dur coup X The boss had a stroke of bad luck 1 2 3 4 5 6
3. Numéroter les segments de gauche à droite d'après leur position (P) et en face de chaque position, indiquer le nombre d'éléments (E) (par exemple de mots) que contient le segment.	P (a) 1 2 3 4 5 6 E (a) 1 1 2 1 1 1 E (x) 1 1 1 3 1 1 P (x) 1 2 3 4 5 6
4. En commençant par la série de l'intérieur re-arranger chaque séquence qui se fait face selon les lignes d'équivalence, faire la soustraction entre le plus grand et le plus petit nombre puis additionner les différences.	E(a) 1 1 2 1 1 1 E(x) 1 1 1 3 1 1 $\Sigma E\|a-x\|$ = 0 +0 +1 +2 +0 +0 = 3 P(a) : 1 2 3 4 5 6 P(x) : 1 2 3 4 5 6 $\Sigma P\|a-x\|$ = 0 +0 +0 +0 +0 +0 = 0

5. Porter le résultat de la différence entre le nombre d'éléments E (a - x) et les positions dans la séquence P (a - x).

$E|a-x| = 3$

$P|a-x| = 0$

6. Prendre l'une après l'autre chaque paire équivalente pour en comparer les constituants formels (c'est à dire les lettres.

Le patron

The boss

7. Découper en graphèmes juxtaposés et faire correspondre les équivalents.

L e p a t r o n

Th e b o s s

8. Faire pour chaque segment le total des graphèmes (e) et faire en nombres réels la soustraction entre les différences de longueur.

		2		6		
		- 3		4		
$\Sigma e\|a-x\|$	=	1	+	2	=	3

9. Additionner tous les graphèmes.

		2		6		
		+ 3		4		
		5		10		

10. Compter les paires assorties.

2 0

11. Faire le total de tous les graphèmes des segments.

$:\Sigma e|a+x|_1$
$+\Sigma e|a+x|_2$

5	+	10	=	15		

12. Faire le total $^{(P)}$ des éléments identiques (≡)	$\Sigma\Theta \lvert a \equiv x \rvert$ = 2 + 0 = 2		
13. Exprimer les totaux sous forme de proportion ou de pourcentage		2/15 = .13	
14. Calculer les différences d'ordre des segments, de longueur des segments, de longueur des mots et de nombre d'éléments des mots	Ordre segmental	0/12	
	Longueur segmentale	3/16	
	Longueur des mots	3/16	
	Eléments des mots	3/15	
15. Exprimer le résultat en pourcentage ou de façon équivalente	Ordre segmental	0	
	Longueur segmentale	0,12	
	Longueur des mots	0,12	
	Eléments des mots	0,13	
16. Faire le total et la moyenne	37/4	0,09.	

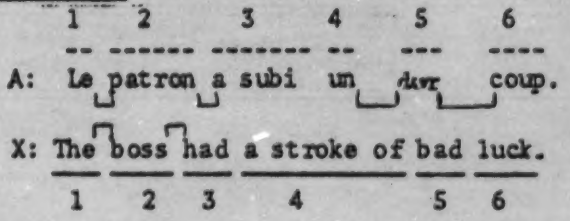
Ce chiffre indique la distance moyenne entre les caractéristiques formelles des deux segments de discours. On remarquera que les mesures sont uniquement fondées sur ce que présentent les textes et ne font pas appel aux codes correspondants.

On peut quantifier des renseignements supplémentaires en ce qui concerne les différences fonctionnelles entre les textes, particulièrement dans la mesure où elles touchent le locuteur, en se référant simplement aux codes respectifs. Il y a cependant un inconvénient à cette procédure: l'analyse peut devenir moins objective et moins précise car elle dépendra alors de la connaissance grammaticale et sémantique que l'on

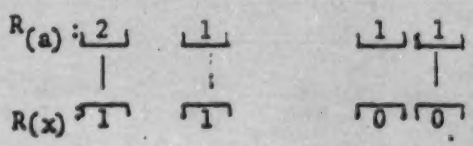
a à propos des opérations du code derrière le message. Voici cependant quelques exemples de développements qui visent à faire ressortir les différences entre les relations et les composantes sémantiques

a) **Relations (R)**

1) Indiquer la dépendance syntaxique à l'aide d'un trait horizontal, de cette façon: $\overline{a_1 \quad a_2}$



2) Numéroter les catégories de relation obligatoires (R), par exemple le genre et le nombre.



3) Redisposer chaque séquence juxtaposée selon les lignes d'équivalences, faire la soustraction entre le plus grand et le plus petit nombre puis additionner les différences.

$R_{(a)}$: 2 1 1 1

$R_{(x)}$: 1 1 0 0

$R|a\text{-}x|$: 1 + 0 + 1 + 1 = 3

b) **Composantes Sémantiques (M)**

1) Compter dans un dictionnaire, le nombre de sens de X correspondant à un élément de A

Anglais	Français
boss	(patron (chef (contremaitre

$M(a \Rightarrow 3x) = 3$

2) Compter dans un dictionnaire, le nombre de sens de A correspondant à un élément de X

Français	Anglais
patron	(boss (master (head (skipper (owner (employer

3) Faire la soustraction entre les deux nombres. $M^{(a \sim x)} = 7 - 3 = 4$

4) Faire le même calcul pour d'autres mots du texte (par exemple the (30) - le (43) = 13 et exprimer le résultat sous forme de proportion par rapport au nombre total de significations.

	The/ Le		boss/patron	
Le =	43	+	7	7
The -	30	+	-3	+3
$M^{(a \sim x)}$ =	13 / 73		4	/10

5) Juxtaposer les distances graphiques (G). G = 3/5 + 10/6 etc. = 13

6) Réduire en pourcentages les deux chiffres.

M =	.19	+	.04
G =	.6	+	1.0

7) Exprimer le résultat sous forme d'une suite de rapports. .19/.6 + .04/1 + ...

On pourrait expérimentalement établir une corrélation entre ces mesures de distance entre langues et les mesures psycholinguistiques d'identification entre langues.

5.0.2 Textes Phonétiques

Quand l'analyse atteint le niveau du mot, on peut se demander si on va utiliser les textes orthographiques ou les textes phonétiques. Les exemples ci-dessus ont utilisé les textes orthographiques. Examinons maintenant ce que ces mêmes textes donneraient si on les analy-

sait comme deux courants de paroles. On peut d'abord découper les textes en syllabes de façon à analyser les différences prosodiques avant de les considérer comme une suite de phonèmes.

5.0.2.1 Segmentation en syllabes

Une langue inconnue peut d'abord frapper par son rythme particulier c'est à dire la façon dont alternent syllabes fortes et syllabes faibles: c'est ce que pensera par exemple un anglophone unilingue en présence du hongrois. Ou alors, la première chose qui frappe sera la montée et la descente de la voix c'est ce que ressentira celui qui ne connaît pas le chinois ou le suédois en écoutant ces langues. Dans les deux cas, la première chose qui attire l'attention n'est pas tant la qualité des sons individuels que le ton et la force des syllabes.

Si l'intonation et le rythme ne diffèrent pas trop de ce à quoi l'auditeur est habitué, il pourra repérer certains mots familiers; par exemple dans un texte hongrois le mot Canada pourra être deux fois plus facilement repérable pour un anglophone que le même mot dans un texte français même si sur la plupart des autres points le français est plus différent du hongrois que de l'anglais (voir ci-dessus).

Les langues peuvent sembler plus différentes par le rythme et l'intonation que par les phonèmes qui les constituent. C'est à cause de ce fait qu'il est parfois pratiquement impossible de reconnaître des homonymes quand ils sont prononcés dans une autre langue. Par exemple, comparons les deux mots suivants:

Fr développement An development

Fr [d e v ə l ɔ p ˈ m ɑ̃] / An [d ɪ ˈ v ɛ l ə p m ə n t]

L'analyse prosodique étudie les faits de rythme et d'intonation; les faits d'enchaînement seront plus commodes à traiter avec l'analyse phonétique. L'unité de l'analyse prosodique est la syllabe. En utilisant la façon de juxtaposer les syllabes qui conviennent le mieux, et en les basant sur les paires lexicales correspondantes déjà dégagées, sur quels points peuvent différer deux syllabes, mis à part les différences entre les phonèmes qui les constituent?

D'un côté la force et la longueur des syllabes peuvent différer et de l'autre la hauteur et la direction de la voix peuvent aussi différer. Des modifications répétées des deux premiers traits produisent des différences de rythme perceptibles; les variations des deux derniers points révèlent des différences d'intonation. Afin d'analyser et de mesurer ces différences, on pourra utiliser une façon simple de noter chaque trait ainsi que l'échelle numérique correspondante (voir ci-dessus § 3.2.2.).

Si l'on applique par exemple cette échelle de notation aux deux phrases déjà étudiées, on pourra mesurer le degré de différence prosodique.

1. Juxtaposer les textes équivalents et les découper en syllabes.

(A) Le patron a subi un dur coup.
(X) The boss had a stroke of bad luck.

2. Faire correspondre les syllabes de la façon la plus appropriée c'est à dire syllabe accentuée en face de syllabe accentuée de façon à ce que les syllabes supplémentaires soient non-accentuées.

(A)	Le	pa	tron	a	su	bi	un			
(X)	The		boss	had			a	stroke	of	
		dur	coup							
		bad	luck							

3. Faire le total des vides dans la ligne inférieure. Additionner les deux sommes. On a la différence de nombre de syllabes.

Le	pa	tron	a	su	bi	un			dur	
The		boss	had			a	stroke	of	bad	
coup	2									
luck	3									

$L(\overline{A} - \overline{X}) = 2$

$L(\overline{X} - \overline{A}) = 3$

$L(\overline{A} \sim \overline{X}) = 5$

4. Pour les syllabes restantes, portez le rythme de la phrase = la force (F) et la longueur (L) et l'intonation: la hauteur (H) et la direction (D) du ton (voir Tableau 6).

A	Le	tron	a	un		dur	coup
X	The	boss	had	a		bad	luck

5. Numéroter les éléments de la ligne a d'après la table des valeurs (voir table)

f	(a)	1	1	1	1	1	2
l	(a)	1	1	1	1	1	2
d	(a)	0	0	0	0	0	0
h	(a)	1	3	1	1	3	0

6. Répéter l'opération pour la ligne x.

f (x) 0 1 0 0 1 2

1 (x) 0 1 0 0 1 2

d (x) 0 0 0 0 0 1

h (x) 0 2 0 0 3 3

7. Faire correspondre chacune des lignes.

1. f (a) - f (x)
2. 1 (a) - 1 (x)
3. d (a) - d (x)
4. h (a) - h (x)

f (a) 1 1 1 1 1 2

f (x) 0 1 0 0 1 2

f|a-x| 1 +0 +1 +1 +0 +0 = 3

f|a-x| = 3

8. Calculer la différence pour chaque ligne.

1 (a) 1 1 1 1 1 2

1 (x) 0 1 0 0 1 2

1|a-x| 1 0 1 1 0 0 = 3

1|a-x| = 3

d (a) 0 0 0 0 0 0

= 1

d (x) 0 0 0 0 0 1

d|a-x| 0 +0 +0 +0 +0 +1 = 1

d(a-x) = 1

h (a) 1 3 1 1 3 0

h (x) 0 2 0 0 3 3

h|a-x| 1 1 1 1 0 3 = 7

h|a-x| = 7

9. Faire le total

$L|\overline{a}\sim\overline{x}|$ = 14

5.0.2.2 Segmentation en phonèmes

Si on pousse l'analyse au-dessous du niveau prosodique, on peut comparer les transcriptions phonétiques de deux textes équivalents pour calculer la distance des constituants et la distance vectorielle entre leurs traits distinctifs. C'est ainsi qu'avec les exemples déjà employés, on procédera de la façon suivante avec les textes en transcription phonétique

Opérations	Exemples
1. Transcrire phonétiquement les phrases équivalentes pour inclure les modifications d'enchaînement.	lə patrɔ̃ a sybi õẽ dy:ʀ ku ðə bɒs həd ə strouk əv bæd lʌk
2. Faire correspondre les mots ou groupes de mots équivalents.	lə patrɔ̃ a sybi õẽ dy:ʀ ku ðə bɒs həd ə strouk əv bæd lʌk
3. En suivant, de gauche à droite, faire correspondre à l'intérieur des segments équivalents les phonèmes de la façon la plus appropriée: les voyelles en face des voyelles, les consonnes en face des consonnes. Porter le signe (/) s'il n'y a pas de phonème correspondant.	lə patrɔ̃ /a sybi õẽ///// // dy:ʀ ku / ðə bɒs// hə //d/ ə strouk əv bæd/ lʌk 2 + 1 + 2 + 1 + 5 + 2 + 1 = 14

4. Compter tous les signes (1). L/a~x) = 14 (Somme des différences de nombre de phonèmes par équivalent sémantique)

5. Indiquer sous chaque paire de phonèmes la distance entre les deux phonèmes (en valeur vectorielle) et additionner toutes les différences (Voir Tableaux 7 et 8.

l e p a t a b œ̃ d y: ʁ k u
o ə b p s ə d ə b ø d l ʌ
3+ 1 +1 +3 +1 +2 +3 +2 +3 +4 +3 +5 +3 = 34
E /a~x) = 34 (Somme des distances entre phonèmes correspondants.

On peut considérer ces textes juxtaposés comme des segments statiques ou comme des séquences dynamiques car, comme nous l'avons vu,il existe deux types de mesure: la mesure otatique et la mesure dynamique qui peuvent chacune s'appliquer de façon différente au-dessous du niveau prosodique où les traits phoniques distinctifs entrent dans les calculs. Une fois que les textes ont ainsi été traités analytiquement, on peut appliquer quelques formules simples pour chaque type de mesure: la mesure statique et la mesure dynamique. Commençons par les mesures statiques.

5.1 Mesures statiques

Ce dont nous avons besoin ici est une formule pour mesurer la distance entre deux textes statiques juxtaposés tels qu'ils apparaissent au regard. Comme ils ne se présenteront pas de la même façon au-dessus et en-dessous du niveau prosodique car dans ce dernier cas interviennent les valeurs des traits distinctifs phoniques, il faudra tenir compte de cette ligne de séparation. Notre premier travail consistera à

élaborer une formule intégrée pour mesurer la distance entre les segments supra-prosodiques.

5.1.1 Segments supra-prosodiques

On devra commencer comme précédemment par traiter le texte: on juxtaposera et numérotera les segments équivalents (N) de gauche à droite puis on en indiquera les constituants. Exemple

	1	2	3	4	5		
A	Il	aime	beaucoup	la bière	de	Munich	N (a) = 5
X	He	likes	Munich	beer	very	much	= 5
	1	2	3	4	5		= 10

Dans le cas présent, on découpe d'après l'équivalence minimale; bien qu'un découpage par couches successives ou par transformations soit également possible, ces procédures risquent d'alourdir inutilement l'analyse et le calcul.

La démarche suivante consistera à calculer le nombre de transpositions (T) nécessaires pour faire concorder les deux structures. Ceci se fait en faisant correspondre les positions équivalentes et en additionnant les différences de position en nombres naturels:

	1	2	3	4	5
A.	Il	aime	beaucoup	la bière	de Munich./
X.	He	likes	very much	beer	Munich.
	1	2	5	4	3
A.	1	2	3	4	5
X.	1	2	5	4	3

T(A-X) = 0 + 0 + 2 + 0 + 2 = 4

Il est ensuite possible de mesurer la différence entre les éléments verbaux constitutifs (V) en comptant le nombre de paires correspondantes équivalentes avec des catégories qui se correspondent terme à terme (C) et en opérant une soustraction entre ce chiffre et le total.

A	Il	aime	beaucoup	la bière	de	Munich /	V (a)	= 7
X	He	likes	very much	beer		Munich	V (x)	= 6
							V (a+x)	= 13
	2	2		2		2	C (a+x)	= 8

On peut exprimer cette différence de configuration des constituants verbaux correspondants par la formule suivante:

$$V(A - X) = V(a + x) - C(a + x)$$

Ce qui donne dans le cas présent $V = V(7 - 6) - C(2222) = 13 - 8 = 5$

Finalement la différence entre les unités (U) qui contiennent ces constituants peut se mesurer en faisant la soustraction entre les paires identiques (p) et le nombre total (n) d'unités après avoir juxtaposé les équivalents de façon à obtenir une correspondance maximale. On peut faire ces opérations sur des textes en transcription orthographique ou en transcription phonologique. Exemple:

b i è r e = 5
b e e r = 4 n = 9

2 + 2 = p - 4

U = n - p = 9 - 4 = 5

/b jɛ:ʁ/ = 4
/bɪəɻ/ = 4 n = 8

2 p = 2

U = n - p = 8 - 2 = 6

M u n i c h = 6
M u n i c h = 6 n = 12

2+2+2+2+2+2 - p = 12

U = n - p = 12 = 12 = 0

/m y n i k/ = 5
/m ju n ɪ k/ = 6 n = 11

2 + 2 + 2 = p = 6

U = n - p = 11 -6 = 5

De cette façon, on peut établir la distance entre homographes et celle entre homophones (la distance entre les textes orthographiques entiers est de 54 - 32 = 22).

Nous avons maintenant trois mesures (U, V et T) des structures et des éléments constituants des textes juxtaposés, on ne peut additionner ces mesures car elles représentent des choses fondamentalement différentes. On doit d'abord les réduire à un dénominateur commun avant de pouvoir les intégrer.

On y arrivera en plaçant dans chaque cas les différences minimum et maximum entre 0 et 1. Si dans chaque cas on a fixé le minimum à 0, il faut alors déterminer quel sera le maximum pour U, V et T. Dans le cas de U, si toutes les formes graphiques juxtaposées sont identiques le maximum de $U = n$ et $\frac{\max (U)}{n} = 1$

Dans le cas de V^n, si tous les constituants équivalents sont compatibles pour ce qui est de la catégorie et du niveau, on aura

$\frac{\max (V)}{v} = 1$ ou $\frac{\max (n - p)}{n} = 1$ et $\frac{\max (v - c)}{v} = 1$

Dans le cas de T si les segments équivalents occupent la même position, alors min (J) = 0; mais comment déterminer le maximum? Si tous les segments juxtaposés sont numérotés comme ci-dessus, en fonction de leur position dans la séquence et si l'on fait la soustraction entre les numéros de position en nombres positifs, on obtiendra une mesure de la différence structurale entre les textes. Quelle est la limite maximum de différence de ce type? Manifestement, on atteint le maximum de diffé-

rence quand les paires d'équivalents se correspondent en ordre inverse.

Exemple

			1		1
Fr		compétence		administrative	
An	administrative			competence	
			2		1

	1	2	3
All	Hier	bin	ich
Fr	Je	suis	ici
	3	2	1

Examinons ces relations et développons-les de façon à en extraire la structure.

1 2 = 2 positions

$$-\frac{2}{1} + \frac{1}{1} = 2 \qquad \frac{2 \times 2}{2} = \frac{2^2}{2}$$

1 2 3 4 = 4 positions

$$\frac{4}{3} + \frac{3}{1} + \frac{2}{1} + \frac{1}{3} = 8 \qquad \frac{4 \times 4}{2} = \frac{4^2}{2}$$

1 2 3 4 5 6 = 6 positions

$$\frac{6}{5} + \frac{5}{3} + \frac{4}{1} + \frac{3}{1} + \frac{2}{3} + \frac{1}{2} = 16 \qquad \frac{6 \times 6}{2} = \frac{6^2}{2}$$

N positions:

$$\frac{N \times N}{2} = \frac{N^2}{2}$$

Remarquons cependant que dans les exemples ci-dessus, N est toujours un nombre pair.

Supposons que le nombre total de positions soit impair (3, 5, 7...). Examinons alors ce qu se passe:

1 1 position $\qquad \frac{(1 \times 1) - 1}{2} = \frac{1^2 - 1}{2}$

$$-\frac{1}{0} = 0$$

1	2	3			3 positions $\frac{(3 \times 3) - 1}{2} = \frac{3^2 - 1}{2}$
3	2	1			
2	0	2	=	4	

1	2	3	4	5	5 positions $\frac{(5 \times 5) \quad 1}{2} = \frac{5^2 - 1}{2}$
5	4	3	2	1	

N positions $\frac{N \times N - 1}{} = \frac{N^2 - 1}{}$

La formule pour le nombre maximum de transpositions structurales dépendra donc du fait que le nombre total de segments est pair ou impair, de telle sorte que si N est pair

$$\text{Max T (N)} = \frac{N^2}{2} \quad \text{et si N est impair max T (N)} = \frac{N^2 - 1}{2}$$

$$\text{Max T (N)} = \begin{cases} \frac{N^2}{2} & \text{si N est pair} \\ \frac{N^2 - 1}{2} & \text{si N est impair} \end{cases}$$

Si bien que le rapport entre le carré du nombre total de segments de deux structures juxtaposées et le double du nombre de transpositions nécessaires pour qu'elles concordent varie entre 0 et 1.

Donc:

$$\max T(N) = \frac{N^2}{2} \quad \text{et} \quad \frac{2}{N^2} \max T(N) = 1$$

$$\max T(N) = \frac{N^2 - 1}{2} \quad \text{et} \quad \frac{2}{N^2 - 1} \max T(N) = 1$$

$$0 \leq \frac{2T}{N^2} \leq 1 \qquad \max\left(\frac{V}{v}\right) = 1$$

$$\max\left(\frac{U}{n}\right) = 1$$

Donc la distance supra-prosodique (D) entre deux structures juxtaposées d'une langue A et d'une langue X est égale au double du nombre de transpositions compensatrices divisé par le carré de la somme des segments équivalents, plus la différence entre les éléments constituants (V) plus le nombre total de composantes graphiques moins le nombre de paires identiques, la valeur de chaque ensemble variant entre 0 et 1.

$$\bar{D}\left(\frac{A}{X}\right) = \frac{2T}{N^2} + \frac{V}{v} + \frac{U}{n}$$

Ces mesures intégrées doivent donc varier entre 0 et 3.

$$0 \leq \bar{D}\left(\frac{A}{X}\right) \leq 3$$

mais $\frac{U}{n} = \frac{n - p}{n}$. Donc la formule intégrée pour les différences statiques entre textes analysés au-dessus du niveau prosodique est:

$$0 \leq \bar{D}\left(\frac{A}{X}\right) \leq 3 = \frac{2T}{N^2} + \frac{u - c}{u} + \frac{n - p}{n}$$

5.1.2 Segments infra-prosodiques

Si les différences de distance entre langues au-dessous du niveau prosodique doivent être prises en compte, on doit mesurer des traits qui sont acoustiques et non plus visuels et les présenter de façon à ce qu'ils traduisent les degrés de ressemblance. Les traits acoustiques traitent des variables prosodiques de la syllabe; les traits allophoniques traitent des différences de prononciation entre phonèmes dans le discours suivi. Ceci implique que pour la formule intégrée on devra remplacer un terme (le graphique - phonologique) par deux (le syllabique et le phonétique).

5.1.2.1 Unités syllabiques

Pour ajouter une mesure de la distance prosodique à la formule, on doit appliquer des procédures de mesure syllabique avant d'arriver à la formule pour l'intégrer.

Comme nous l'avons déjà expliqué, les opérations de mesure prosodique sont les suivantes: on juxtapose les syllabes des segments de discours équivalents de façon à obtenir une correspondance optimale; on note, pour chaque syllable, le rythme (force et longueur) et l'intonation (hauteur et direction), puis on mesure chacune de ces quatre composantes phoniques sur une échelle de valeurs relatives (voir ci-dessus § 4.2.0).

Par exemple, prenons l'homographe Munich tel qu'il est prononcé en anglais et en français dans ces contextes comparables. On

commence par découper ce segment de discours en syllabes et pour chaque syllabe, on note la longueur, la force, la hauteur et la direction relatives:

An	\ mju	nĭk	=	h	3	2	f	1	0
				d	1	2	1	0	0
Fr	| my	nik	=	h	3	1	f	2	2
				d	0	1	1	0	1

On donne les valeurs numériques qui conviennent aux différences et on traite chacune d'elles comme un vecteur prosodique en effectuant entre elles des soustractions en nombres entiers positifs. Exemple:

		f	l	d	h	
An	\ mju	2	2	1	3	
Fr	| my	1	1	0	3	
		1	+1	+1	+0	= 3
An	nĭk	0	0	0	1	
Fr	| nik	2	1	0	2	
	D =	2	+1	+0	+1	= 4/7

Enfin on additionne les différences vectorielles pour chaque paire de syllabes (3 + 4) afin d'obtenir la distance prosodique (7).

Pour résumer, on calcule la différence prosodique entre deux segments de discours équivalents en prenant la somme (Σ) des différences entre chaque syllabe juxtaposée (S), en chaque position (q) jusqu'à ce qu'on ait épuisé toutes les positions syllabiques (Q); ce qui donne:

$$\sum_{q=1}^{Q} S_q$$

Pour l'exemple ci-dessus on aurait les valeurs suivantes: Q = 2 positions syllabiques, S_1 = 3 et S_2 = 4

$$\sum_{q=1}^{Q} = 3 + 4 \qquad \sum_{q=1}^{Q} S_q = 7$$

Le problème qui se pose ensuite est d'intégrer ceci dans la formule générale. Cela peut se faire en exprimant la distance syllabique sous la forme d'un pourcentage de la limite (L) du maximum (max) qui est de 12.

TABLEAU 11

Signes et Symboles pour les mesures intégrales

Q : nombre total des syllabes juxtaposées (y compris les cases vides)	K : nombre total des phonèmes juxtaposés (y compris les cases vides)
q : une position syllabique donnée	k : une position de phonème donné
Sq:: différence syllabique à la position q	P_k:: différence phonémique à la position k
L : limite prosodique différentielle supérieure	B : limite phonémique différentielle supérieure
l : limite prosodique différentielle inférieure	b : limite phonétique différentielle inférieure
U : différence d'unités	u : nombre d'unités
V : différence de constituants	v : nombre de constituants
C : catégorie commune	c : nombre de constituants communs
N : nombre de segments	n : nombre d'éléments
T : transposition	p : nombre de paires assorties

$$\max\left(\sum_{q=1}^{Q} S_q\right) = \sum_{q=1}^{Q} \max S_q = BQ$$

Cela signifie que la différence maximale entre les traits prosodiques de deux segments de discours quelconques (LQ) est égale à la limite différentielle (L) multipliée par le nombre total de syllabes, soit la différence maximum entre deux syllabes. Si l'on veut intégrer la formule pour la différence prosodique, on doit donner 1 comme valeur maximum et l'on aura alors:

$$\frac{1}{LQ} \max\left(\sum_{q=1}^{Q} S_q\right) = 1$$

On devra ajouter ceci à la formule dynamique infra-prosodique en tant que

terme nouveau exprimant les différences prosodiques de prononciation de deux segments de discours quelconque comme étant égales à la somme des différences entre toutes les syllabes juxtaposés $\left(\sum_{q=1}^{Q} S_q\right)$ divisée par le nombre total de syllabes (Q) multiplié par la limite prosodique différentielle, ce qui donne $\frac{\sum_{q=1}^{Q} S_q}{L\ Q}$

Comme la position dans une phrase est parfois fonction de l'intonation et du rythme, il est peut-être utile de répéter les calculs en juxtaposant des phrases entières comme des suites de syllabes.

5.1.2.2 Unités phonémiques

On peut mesurer avec plus de précision encore la distance entre les constituants phonétiques des syllabes. C'est à ce niveau que l'on tiendra compte des effets d'enchaînement souvent analysés en même temps que les constituants prosodiques du discours. Ainsi, on pourrait transcrire, découper et juxtaposer de la façon suivante l'homographe français-anglais: observation

Fr	ɔ	p	s	ɛ	ʁ	v	a	s	j	ɔ̃	/
An	ɒ	b	z	ə		v	e	ʃ	ə	ŋ	

Avant d'entreprendre toute mesure, il faut transcrire les deux textes juxtaposés de façon adéquate. On indiquera sous chaque paire correspondante la différence vectorielle. Pour l'exemple proposé, on utilisera la table des distances entre phonèmes (Table 1). Ensuite, on additionnera toutes

les différences suivant la formule suivante: $\sum_{k=1}^{k} P_k$ dans laquelle K représente le nombre total de positions, k une position donnée quelconque et P_k la distance vectorielle entre deux phonèmes quelconques se correspondent en position K. A titre de comparaison, reprenons un exemple déjà donné:

Fr bière / An beer

k_1 k_2 k_3 k_4 = 4 = K

Fr b j ɛ: ʁ

An b I ə ɹ

0 + 2 + 2 +4 = 8

$$\sum P_k = P_1 + P_2 + P_3 + P_4 = \sum_{k=1}^{k} P_k = 8$$

Afin d'intégrer cette mesure dans la formule générale, il faut que tous les résultats possibles soient compris entre 0 et 1. Cela signifie qu'il faut que les calculs soient effectués entre certaines limites: une limite inférieure (b) et une limite supérieure (B). Si deux phonèmes sont phonétiquement identiques comme /m/ en français et en anglais, la distance vectorielle entre la façon de les prononcer est 0 qui constitue la valeur la plus faible ou la limite inférieure (b) et (b) = 0. Les deux phonèmes qui dans une juxtaposition des codes phonétiques diffèreront le plus représenteront la limite supérieure de la différence possible entre un phonème d'une langue et un phonème de l'autre langue. D'après notre table, on aurait une valeur de 17,5 pour l'anglais et le français, mais comme il peut y avoir des différences de longueur conditionnées par le contexte dans des segments d'énoncés réels, nous ajouterons un point supplémentaire pour tenir compte

de toutes les différences de cet ordre si bien que B / Fr ~ An / = 18,5 .

La différence maximum (BK) entre les traits phonétiques de deux segments équivalents de discours quelconques est égale à sa longueur dans le nombre total de positions (K) multipliée par la différence maximum (max P_k) entre deux phonèmes quelconques pris chacun dans l'une des langues que l'on veut comparer, si bien que:

$$BK = \sum_{k=1}^{K} \max P_k$$

Pour pouvoir être intégré dans notre formule, il faut que ce maximum soit égal à 1, ce qui donnera

$$\frac{1}{BK} \max \sum_{k=1}^{k} P_k = 1$$

La différence phonétique entre la prononciation de deux segments de discours quelconques est égale au nombre total de phonèmes juxtaposés $\sum_{k=1}^{k}$ (y compris les cases vides) que multiplie leurs différences k=1 juxtaposées (P_k) divisé par leurs différences vectorielles maximum que multiplie le nombre de positions (K), ce qui donne: $\frac{\sum_{k=1}^{K} P_k}{BK}$. Si l'on applique cette formule à l'exemple ci-dessus où K = 4, P_k = 8, B = 18,5 , nous aurons

$$\frac{\sum_{k=1}^{K} P}{BK} = \frac{8}{18.5 \times 4} = \frac{8}{74} = .108$$

La formule générale développée en-dessous du niveau du phonème implique par conséquent les quatre termes suivants:

1. $0 \leq \left(\frac{2T}{N^2} \text{ or } \frac{2T}{N^2 - 1} \right) \leq 1$

2. $0 \leq \left(\frac{u - c}{u} \right) \leq 1$

3. $\frac{1}{L} \leq \left(\frac{\sum_{x:T} S_q}{L Q} \right) \leq 1$

4. $\frac{b}{B} \leq \left(\frac{\sum_{x:T} P_x}{B K} \right) \leq 1$

Comme chacune de ces formules varie entre 0 et 1, la formule intégrée variera entre 0 et 4. Voici la formule intégrée pour la distance statique développée en-dessous du niveau du phonème $\left(\frac{A}{X} \right]$ entre deux segments équivalents de discours quelconques:

$$0 \leq \bar{D}\left(\frac{A}{X} \right] \leq 4 = \frac{2T^*}{N^2} + \frac{u - c}{u} + \frac{\sum_{x:T} S_q}{L Q} + \frac{\sum_{x:T} P_x}{B K}$$

*$\left(\frac{2T}{N^2 - 1} \right)$ Si on a un chiffre impair.

5.2 Mesures dynamiques

Pour effectuer les mesures intégrales de distance entre discours, on a considéré les textes juxtaposés comme deux lignes dans l'espace découpées de façon à comparer et à mesurer les équivalents correspondants. Dans ce type de mesure statique, on n'a pas tenu compte de l'élément temps de la communication.

Si l'on tient compte de l'élément temps ou plus précisément de l'ordre d'apparition des unités constituantes, les segments deviennent des séquences avec les caractéristiques que cela entraîne quand on veut trouver la différence entre deux séquences de discours correspondantes, on peut mesurer les constituants juxtaposés en se plaçant de deux points de vue: soit ils sont identiques ou différents, soit ils diffèrent de degré.

Dans le premier cas, comme on se trouve au-dessus du niveau prosodique (: supra-prosodique), c'est par les représentations graphiques du texte (qu'elles soient orthographiques ou phonologiques) que l'on découvrira les différences. Comme nous l'avons vu, cette représentation graphique du discours peut se limiter à la suite des phonèmes ou se pr présenter sous la forme d'une transcription très détaillée de sons.

5.2.1 Séquences supra-prosodiques

Les textes juxtaposés se présentent avec des éléments disposés dans un certain ordre et comme on peut traiter l'ordre spatial comme étant analogue à l'ordre temporel, on peut utiliser la même formule de transposition qui a été établie pour mesurer la distance entre structures, c'est à dire $\frac{2\ T}{N^2}$.

Ce qui nous intéresse maintenant, ce n'est pas seulement de savoir quels sont les éléments équivalents qui sont identiques, mais aussi en quel endroit ou plus exactement à quel moment. En d'autres termes: à quel point de la séquence temporelle l'identité apparaît-elle? Cela signifie qu'il faut tenir compte non seulement du nombre de positions mais de

l'ordre dans lequel elles sont placées dans la séquence.

Par exemple, mesurons la distance dynamique entre le mot français formules et l'équivalent anglais (nord-américain) formulas, ce qui donne en transcription phonologique:

Fr f ɔ ʁ m y l (a)

An f ɔ ɹ m ju l ə z (x)

Quand on examine les transcriptions juxtaposées ci-dessus, on peut faire les remarques suivantes:

1- La forme française occupe six positions alors que l'équivalent anglais en occupe neuf.

2- Sur le maximum de neuf positions, six ont des équivalents correspondants et trois, /j, ə, z/, n'en ont pas. Nous désignerons ce groupe par le symbole K : K = {j, ə, z}.

3- Parmi les positions qui ont deux phonèmes juxtaposés, certains ont deux phonèmes identiques, d'autres deux phonèmes différents. Nous désignerons ce dernier groupe par le symbole L : $L = \left\{\begin{matrix} R & y \\ ɹ & u \end{matrix}\right\}$.

4- Quatre des phonèmes sont identiques dans les deux segments de discours: c'est le groupe {f, ɔ, m, l} qui est manifestement égal à zéro: {f, ɔ, m, l} = 0 . Il faut maintenant mesurer les différences.

5- Nous remarquons que R/ɹ occupent la position 3 et y/u la position 6 ce qu'on peut exprimer par L{3,6} ou L = $\{L_3, L_6\}$.

6- On peut exprimer sous la forme d'une moyenne chacune des positions des différentes unités correspondantes (a + x), c'est à dire (6 + 9) ce qui donne $\frac{a + x}{2} = \frac{6 + 9}{2} = \frac{15}{2}$

7- On peut exprimer la valeur du dénominateur commun par un facteur équivalent au nombre de paires correspondantes (2) de la séquence c'est à dire $(1_3, 1_6)$ en les considérant comme des exposants dont on a réduit les valeurs à des puissances de 2 par exemple (3 - 2, 6 - 2) ce qui nous donne 2^{3-2}, 2^{6-2} ou de façon plus générale 2^{L-2}. En utilisant ce dénominateur variable pour chaque paire non-identique, on obtient des valeurs qui décroissent en fonction de la distance entre l'unité et la première position, c'est à dire le début du mot dans le cas présent. Ce qui donne $R/d = 15/2^{3-2}$ et $y/u = 15/2^{6-2}$ ou de façon plus générale $\frac{a + x}{2^{L-2}}$

8- Quand on examine le premier groupe de mots (∅, x) qui n'ont pas de correspondants {j, ə, z}, c'est à dire le groupe K qui occupent les positions $\{k_5, k_8, k_9\}$, il nous suffira pour adapter la formule de réduire d'un la puissance de l'exposant de façon à ce que 2^{k-1} devienne le dénominateur. Par conséquent $\{\emptyset/j\} = 2^{5-1}$; $\{\emptyset/ə\} = 2^{9-1}$ ou de façon plus générale $\frac{a + x}{2^{k-1}}$

9- La formule générale pour mesurer la distance dynamique (W) entre deux séquences équivalentes juxtaposées sera donc:

$$W = \sum_{k \in K} \frac{a + x}{2^{k-1}} + \sum_{l \in L} \frac{a + x}{2^{l-2}}$$

ce qui pour notre exemple donnera

$$W(\text{famyl}/\text{famjulaz}) = \frac{6 + 9}{2^{5-1}} + \frac{6 + 9}{2^{8-1}} + \frac{6 + 9}{2^{9-1}} + \frac{6 + 9}{2^{3-2}} + \frac{6 + 9}{2^{6-2}} =$$

$$\frac{15}{16} + \frac{15}{128} + \frac{15}{256} + \frac{15}{2} + \frac{15}{16} = 9.106$$

Pour avoir une idée de la façon dont les distances obtenues par ces mesures diffèrent de celles qui sont obtenues par les mesures statiques et aussi des différences entre segments de discours en transcriptions orthographiques et phonétiques, appliquons cette formule aux exemples déjà étudiés: Fr <u>bière</u> et An <u>beer</u>:

$$\begin{array}{lccccccc} & & l_1 & l_2 & l_3 & l_4 & l_5 & \\ a = & & b & i & è & r & e & = 5L \\ x = & & b & e & e & r & & = 4L \end{array} \qquad a + x = 5 + 4 = 9$$

$$K = \{e\} = 1^{k^5} \qquad L = \{i/e,\ e/e\} = \{L_2,\ L_3\}$$

$$K = \{e/\emptyset\} = 2^{5-1} \qquad L = \{2\}$$

$$\sum_{k \in K} \frac{a+x}{2^{k-1}} = \frac{5+4}{2^{5-1}}, \quad \sum_{l \in L} \frac{a+x}{2^{-2}} = \frac{5+4}{2^{2-2}} + \frac{5+4}{2^{3-2}}$$

$$\sum_{k \in K} \frac{a+x}{2^{k-1}} + \sum_{l \in L} \frac{a+x}{2^{l-2}} = \frac{9}{2^4} + \frac{9}{2^0} + \frac{9}{2^1} = \frac{9}{16} + \frac{9}{1} + \frac{9}{2} =$$

$$= .56 + 9.00 + .45 = 10.01$$

$$\underline{W}\left(\frac{\text{bière}}{\text{beer}}\right) = 10.01$$

La même formule appliquée à une transcription phonétique (restreinte) de ces deux mots ferait apparaître une différence plus grande, entraînée par les prononciations différentes du (r). On aurait (bjɛːʁ / bɪəɹ)

$$\begin{array}{cccccc} & L_1 & L_2 & L_3 & L_4 & \\ a = & b & j & \varepsilon: & ɹ & = 4 \\ x = & b & ɪ & ə & ɹ & = 4 \end{array} \qquad a + x = 8$$

$$K\{\emptyset\} = 0 \qquad L = \{l_2, l_3, l_4\} = 3$$

$$\sum_{k \in K} \frac{a+x}{2^{k-1}} + \sum_{l \in L} \frac{a+x}{2^{l-2}} = 0 + \frac{4+4}{2^{2-2}} + \frac{4+4}{2^{3-2}} + \frac{4+4}{2^{4-2}} =$$

$$\frac{8}{2^0} + \frac{8}{2^1} + \frac{8}{2^2} = \frac{8}{1} + \frac{8}{2} + \frac{8}{4} = 8+4+2 = 14. \quad W/\underline{bj\varepsilon:ɹ}/\underline{bɪəɹ} = 14$$

Pour trouver les différences dynamiques entre deux séquences de mots, on devra simplement additionner les différences entre chaque paire correspondante. La différence pour un texte de (I) paires est la somme des différences individuelles, c'est à dire:

$$\sum_{i=1}^{I} W_i = \sum_{i=1}^{I} \sum_{k \in K} \frac{a_i + x_i}{2^{k-1}} + \sum_{i=1}^{I} \sum_{l \in L} \frac{a_i + x_i}{2^{l-2}}$$

5.2.2 Séquences infra-prosodiques

Même si la transcription des segments de discours juxtaposés est très précise, il est possible que le nombre de détails nécessaires pour différencier les sons ne soit pas suffisant. Ceci a des chances de se produire en particulier quand on mesure des dialectes très voisins ou quand on étudie les interférences bilingues. Au-delà du niveau de différenciation graphique, on peut découper les sons en traits distinctifs que l'on place tous sur une échelle comme on l'a fait pour les mesures statiques infra-prosodiques (voir Tableau 6, 7 et 8).

Cependant, maintenant la différence réside dans le fait que la valeur vectorielle réelle de la distance entre deux sons juxtaposés dépend de leur position dans la séquence, ce qu'on a fixé comme équivalence minimale. Comme pour la mesure dynamique de segments juxtaposés de symboles graphiques (orthographiques ou phonétiques), les valeurs attribuées aux différences sont décroissantes: elles diminuent au fur et à mesure que la séquence progresse dans le temps. Toutefois il faut ici traiter les différences, non pas en termes de "identique ou différent" mais en termes de degré de ressemblance. On arrive à ce résultat en faisant des soustractions entre traits distinctifs.

On traitera donc différemment ces deux types de mesure. On n'a plus besoin de tenir compte du fait que les symboles sont identiques, différents ou n'ont pas de correspondant, puisque ceux qui représentent des sons ayant des traits identiques auront une distance de zéro, ceux qui diffèrent seront affectés d'une valeur vectorielle positive et ceux qui n'ont pas de correspondants pourront être affectés soit de la valeur obtenue en faisant la somme de leurs traits distinctifs, soit du maximum divisé par deux. En conséquence, les séquences de sons se présenteront sous la forme de suites de nombres, chaque nombre représentant le degré de différence entre les sons, en fonction de la position de chaque son. La position affecte chaque nombre d'une valeur décroissante qui modifie les degrés de différence. Ensuite, on additionne toutes les valeurs et on en fait la moyenne.

De façon à faire décroître successivement les valeurs des distances vectorielles, on doit prendre la puissance de la position (K) des paires juxtaposées (p), puissances qui augmentent avec la longueur de

la séquence (2^1 , 2^2 , 2^3 , 2^4 ... 2^k = 2, 4, 8, 16 ... n) comme diviseur de ces distances vectorielles qui deviennent alors dividendes (P^k / 2^k). Comme la différence entre deux sons correspondants est aussi fonction de la longueur (K) de leurs séquences équivalentes (i), on doit aussi faire entrer cet élément en ligne de compte: exemple le <u>O</u> de <u>pop</u> / <u>pop</u> (2^3) ne peut recevoir la même valeur que le <u>O</u> de <u>population</u> / <u>population</u> (2^{10}). On peut faire intervenir cet élément comme multiple constant de chaque distance et aussi comme diviseur du total. Voyons comment cela peut se faire pratiquement: pour faciliter les comparaisons, reprenons un des exemples déjà étudiés. On mesurera la différence phonétique entre la prononciation du français <u>bière</u> et de l'anglais <u>beer</u> de la façon suivante:

		k_1	k_2	k_3	k_4
1-	Juxtaposer les phonèmes de façon à obtenir la correspondance séquentielle optimale.	b $\frac{b}{0}$	j $\frac{r}{2}$	ɛ: $\frac{ə}{2}$	ʁ $\frac{1}{4}$
2-	Porter la distance vectorielle (P^k) (voir la table pour l'anglais et le français).	0	2	2	4
3-	Multiplier chaque nombre par 2 élevé à la même puissance que la longueur de la séquence (K_1), dans le cas présent: K_1 = 4 c'est à dire que 2 $2^{ki} = 2^4$	k_1 0.2^4 0	k_2 2.2^4 32	k_3 2.2^4 32	k_4 2.2^4 64

4- Diviser chacun de ces produits par 2 élevé à la puissance donnée par le nombre ordinal représentant la position (2^k).

$$2 \quad 4 \quad 9 \quad 16$$
$$0 + 3 + 4 + 4 = 64$$
$$\frac{64}{16} = 4$$

5- Faire le total et diviser par la longueur des séquences prise comme puissance de la paire ($2^{ki} = 4 = 16$)

$$\frac{0}{2^4} \quad \frac{32}{2^2} \quad \frac{32}{2^3} \quad \frac{64}{2^4}$$

La distance vectorielle dynamique entre Fr /bjɛ:ʁ/ et An /bɪəɹ/ sera donc de 4. On peut résumer l'opération qui consiste à mesurer la distance vectorielle dynamique entre les réalisations phonétiques (P) de deux séquences équivalentes (i) par la formule suivante:

$$D_{P_i} = \sum_{k=1}^{k=n} \frac{D_k}{2^{ki}}$$

Bien que l'on aura peut être pas besoin de mesures aussi fines pour comparer des textes assez longs dans deux langues différentes, une telle formule se révèlera utile pour calculer de façon précise la distance entre deux dialectes ou pour étudier les interférences bilingues. Pour illustrer ceci, appliquons la formule à la transcription d'un mot extrait d'un enregistrement au cours duquel un Acadien bilingue racontait une histoire en

anglais et utilisait des noms de lieux français. L'un de ces noms est Rivière-du-Loup que le locuteur prononce [rivjɛrdy'lu] dans son propre dialecte français, mais qui dans un contexte anglais devient ['ɹivəɹdə'lu]

A (An $_i$) =	ɹ	i	v		ə	ɹ	d	ə	l	u
X (Fr $_i$) =	r	ɪ	v	j	ɛ	r	d	y	l	u
P_x =	6	1	0	12	1	6	.5	3	1	.5

$$P_i = \frac{\frac{6}{2^1} + \frac{1}{2^2} + \frac{0}{2^3} + \frac{12}{2^4} + \frac{1}{2^5} + \frac{6}{2^6} + \frac{.5}{2^7} + \frac{3}{2^8} + \frac{1}{2^9} + \frac{.5}{2^{10}}}{2^{10}}$$

$$\underline{P}_i = \frac{3072 + 256 + 0 + 768 + 32 + 96 + 4 + 12 + .5}{1024}$$

$$\underline{P}_i = \frac{4242.5}{1024}$$

$$\underline{P}_i = 4.4$$

Ceci donne une mesure appliquable à toute séquence de discours juxtaposé équivalente. Pour l'appliquer à un texte (par exemple une phrase ou une série de phrases) contenant un nombre (I) de segments consécutifs, juxtaposés après avoir été transposés de façon à obtenir l'équivalence minimale, on aura besoin d'une formule globale pour additionner les distances entre les séquences correspondantes dont le texte se compose, c'est à dire $\sum_{i=1}^{I}$ si bien que $\sum_{i=1}^{I} \underline{P}_i = \sum_{i=1}^{I} \sum_{n=1}^{K_i} \frac{P_n}{2^n}$

5.2.3 Intégration

Les formules dynamiques ci-dessus mesurent la distance entre deux segments de discours minimaux équivalents au fur et à mesure que chacun de leurs constituants (phonèmes ou lettres) apparaît dans le temps. Mais comment tenir compte de l'ordre de ces segments? Comme on peut les considérer comme analogues aux segments statiques, on peut leur appliquer la même formule $\frac{2T}{N^2}$ entre zéro et un, il faut adapter les formules dynamiques pour qu'elles aussi varient entre zéro et un. Comment procéder? Commençons par intégrer les mesures phonétiques.

Comme les limites supérieures et inférieures (B/b) ne changent pas, on peut pour chaque segment les multiplier par un moins un que divise 2 élevé à la puissance représentée par la longueur (ce chiffre est indiqué parle dernier nombre ordinal de la suite de positions (K_1), c'est à dire $\left(1 - \frac{1}{2^{ki}}\right)$ pour le nombre total de segments de discours $\sum_{i=1}^{I}$ et $\sum_{i=1}^{I}\left(1 - \frac{1}{2^{Ki}}\right)$ ce qui donne:

$$\frac{\max\left(\sum_{i=1}^{I}\sum_{k=1}^{K_i}\frac{P_{ik}}{2^k}\right)}{B\sum_{i=1}^{I}\left(1-\frac{1}{2^{K_i}}\right)} = 1$$

La formule dynamique intégrée pour mesurer la distance vectorielle entre deux textes oraux ($\left|\frac{A}{X}\right|$ sera donc:

$$\overline{\overline{D}}\left[\frac{A}{X}\right] = \frac{2T}{N^2} + \frac{\sum_{i=1}^{I} P_{ik}}{B\sum_{i=1}^{I}\left(1-\frac{1}{2^{K_i}}\right)}$$

$\times \frac{2T}{N^2-1}$ si N est un nombre impair.

184

Finalement, établissons la formule pour intégrer la distance dynamique entre deux textes graphiques. Si on considère les textes comme des séquences graphiques, la limite inférieure est zéro car

$$(W_i) = \frac{1}{\sum_{i=1}^{I}} (\min W_i) = 0 \quad \text{si} \quad K_i = L_i \{\emptyset\}_i$$

Pour la limite supérieure tous les éléments juxtaposés doivent différer $L_i \{1 \ldots a_i = X_i\}$ Ce qui donne

$$\max (W_i) = \sum_{\lambda=1}^{X_i} \frac{X_i}{2^{I-\lambda}} = X_i \sum_{\lambda=1}^{X_i} \frac{1}{2^{I-\lambda}} = X_i \left[8 - 2^{3-X_i}\right]$$

$$\max \left(\sum_{\lambda=1}^{I} W_\lambda\right) = \sum_{\lambda=1}^{I} (\max W_\lambda) = \sum_{\lambda=1}^{I} X_\lambda \left[8 - 2^{3-X_\lambda}\right]$$

$$\max \left(\sum_{\lambda=1}^{I} W_\lambda\right) \Big/ \sum_{\lambda=1}^{I} X_\lambda \left[8 - 2^{3-X_\lambda}\right] = 1$$

La formule dynamique intégrée pour mesurer la distance entre deux textes graphiques sera donc

$$\overline{D}\left(\underset{\sim}{\overset{A}{X}}\right) = \frac{2T}{N^2}{}^{*} + \sum_{\lambda=1}^{I} W_\lambda \Big/ \sum_{\lambda=1}^{I} X_\lambda \left[8 - 2^{3-X_\lambda}\right]$$

* $\frac{2T}{N^2-1}$ si N est un nombre impair

Cette dernière formule est bien sûr moins importante que celle qui sert à mesurer la distance dynamique entre textes oraux parce que l'élément temps affecte moins le mot écrit. Cependant la position d'un élément dans une séquence peut constituer le facteur décisif dans les cas d'identification et d'interférences entre langues car la parole est essentiellement un enchaînement d'associations. C'est pour étudier des questions de ce type qu'il peut être intéressant de mesurer la distance entre langues.

C O N C L U S I O N

Mesurer la distance entre langues est à la fois possible et réalisable et l'on peut opérer de différentes façons. On peut mesurer à l'aide d'une certaine technique la différence entre les codes de deux langues et à l'aide d'une autre technique les différences entre des échantillons de discours. On peut mesurer ces échantillons comme des entités statiques occupant un certain espace ou comme des séquences cinétiques ou dynamiques se déroulant dans le temps. On peut mesurer la distance entre deux langues sous la forme d'une somme des différences entre elles deux ou sous la forme de la quantité de travail nécessaire pour passer d'une langue à l'autre par transformations. Chacune de ces deux mesures peut s'effectuer soit en prenant toutes les caractéristiques par lesquelles les deux langues peuvent différer et en comptant pour chaque caractéristique le nombre de différences soit en intégrant les différences immédiatement observables en une formule unique de mesure.

On peut séparer la mesure des différences entre les organisations sémantiques des langues et la mesure des différences entre leurs traits formels, ou bien on peut considérer que l'équivalence de sens est une constante des textes juxtaposés que l'on compare. On peut aussi mesurer la distance entre langues indirectement en comparant chaque langue à une forme standard ou en comparant les deux langues à une troisième langue.

Le type de mesure à utiliser dépendra du but que l'on s'est fixé et des raisons que l'on a de vouloir mesurer la distance entre deux langues. Le but peut être pratique ou théorique. Les buts pratiques s'apparentent beaucoup aux besoins de mesures de toutes sortes que nous éprouvons dans la vie de tous les jours: mesure du temps, de la température, du poids, du volume et de la vitesse. L'utilité de ces mesures dépend de leur adéquation au but visé. Un thermomètre ordinaire que l'on place dehors suffira par exemple à indiquer si l'on doit ou non se vêtir chaudement pour aller travailler mais il ne suffira pas à indiquer si un enfant qui a de la fièvre peut ou non aller à l'école. Que la température extérieure ou la température du corps soit mesurée en degrés centigrades ou en degrés fahrenheit n'a aucune importance, le tout étant de savoir quelle échelle on utilise.

Quand on étudie de façon pratique les problèmes qui touchent à la langue, des recherches très fines, du type de celles que l'on fait en laboratoire impliqueront des types de mesure différents de ceux qui seront nécessaires pour des recherches plus grossières. L'étude de l'acquisition des langues exigera de même des types de mesure différents de ceux qui conviendront à l'étude des interférences bilingues ou bi-dialectales.

Les effets produits par le contact de deux langues très rsssemblantes seront peut-être très différents de ceux produits par le contact de deux langues dissemblables. Pour pouvoir mesurer la distribution relative de chaque langue dans le comportement d'un individu bilingue, on doit d'abord être capable de distinguer les deux langues. A des

fins d'analyse et de mesure, il ne suffit pas de savoir quelles sont ces langues, il faut aussi connaître le ou les dialectes particuliers employés et le degré de différence entre eux.

Sur un plan théorique les mesures entre langues peuvent également contribuer à résoudre des problèmes généraux en linguistique et en psychologie. Comme un grand nombre de ces problèmes sont des problèmes de catégorisation qui se posent en termes d'identique ou de différent, de plus ou de moins, on pourrait peut-être les résoudre en remplaçant les catégories qui sont actuellement employées par des échelles, c'est à dire en rendant mesurables les différences.

Comme nous l'avons vu, un de ces problèmes est posé par l'existence de catégories linguistiques universelles, grammaticales, lexicales et phonétiques. Est-ce que tous les êtres humains ont en commun certaines caractéristiques relatives à l'esprit, au corps et au milieu, caractéristiques qui créeraient des catégories linguistiques universelles dans leurs codes, leurs discours et leurs systèmes de communication? Au-delà du manifeste et de l'axiomatique de la logique, comment peut-on prouver qu'une catégorie linguistique est universelle? Au cours du processus de vérification, il deviendra peut-être évident que certaines catégories linguistiques sont plus générales que d'autres et que le degré de généralité s'appliquera peut être à des catégories plus importantes que leur proposition universelle: l'alternative "ou / ... ou bien". Autrement dit, la linguistique ne sera peut-être pas en mesure de résoudre le problème des universaux linguistiques tant qu'elle ne sera pas capable de s'exprimer par le langage le plus universel qui soit, le langage des nombres.

QUANTITATIVE INTERLINGUISTICS

A Tentative Bibliography

INTERLINGUISTIQUE QUANTITATIVE

Une bibliographie intérimaire

1. AFENDRAS, Evangelos A., 1968. The Balkans as a Linguistic Area: a Study in Phonological Convergence. Ph.D. Dissertation, The Johns Hopkins University, Baltimore.

2. _____ TZANNES, N. and TREPANIER, J.-G., 1970. The Dynamics of a Pattern Change: Ordered Rules and Feature Entropies. Finite String. VII, 6:9.

3. ALATIS, James, ed., 1968. Georgetown Monograph Series on Languages and Linguistics. Washington: Georgetown University Press, No. 21.

4. ANDREYEV, N.D., ed., 1965. Statistiko-Kombinatornoje modelirovanije jazykov. Moscow.

5. ANDRYUSHENKO, V.M., 1967. Castotnyje Slovari i ikh Parametry. Inostrannye Jazyki v Skole, 3:34-42.

6. ANTSIFEROVA, G.M., 1969. Opyt issledovaniya semanticheskoi mikrostruktury. Filologicheskie Nauki, XII, 5:78-86.

7. AUSTIN, William M., 1957. Criteria for phonetic similarity. Language, XXXIII:538-543.

8. AVRAM, A., 1966. La classification des phonèmes selon leur degré de parenté. Acta Linguistica Hafniensia, IX, 2:173-178.

9. AVRAM, A., 1964. Sur la typologie phonologique quantitative. Revue Roumaine de Linguistique, IX:131-134.

10. BAILEY, R.W. and DOLEZEL, L., 1968. An Annotated Bibliography of Statistical Stylistics. Ann Arbor: University of Michigan, Department of Slavic Languages.

11. BENDIX, E.H., 1966. Componential Analysis of General Vocabulary: The Semantic Structure of a Set of Verbs in English, Hindi and Japanese. Den Haag/Paris: Mouton.

12. BENVENISTE, E., 1966. Convergences Typologiques. Homme, VI, 2:5-12.

13. BIRNBAUM, Henrik, 1966. On Typology, Affinity and Balkan Linguistics. Zbornik za filologiju i lingvistiku, IX:17-30.

14. BOOTH, K.H.V., 1970. Statistical parsing by computer. Pensiero e Linguaggio in Operazioni, I, 1:107-110.

15. BOSAK, J., 1970. Review: A Comparative Quantitative Phonology of Russian, Czech and German, by KUCERA, H. and MONROE, G.K. Jazykovedny Casopis, XXI, 1:110-112.

16. BUJAS, Zeljka, 1969. Computers in the Yugoslav Serbo-Croat/English Contrastive Analysis Project. International Conference on Computational Linguistics (Preprint 26) Stockholm: KVAL.

17. CARROLL, John B., 1958. The assessment of phoneme cluster frequencies. Language, XXXIV:267-278.

18. CECCATO, S., 1970. Linguistica operativa e progetto di lingua universale. Pensiero e Linguaggio in Operazioni, I, 1:3-19.

19. _____ Operational Linguistics and Design of Universal Language. Pensiero e Linguaggio in Operazioni, I, 1:20-37.

20. CHRETIEN, C. Douglas, 1962. The mathematical models of glottochronology. Language, XXXVIII:11-37.

21. DARBELNET, Jean, 1967. Composantes sémantiques. La Revue canadienne de linguistique, XIII:15-19.

22. _____ 1971. Linguistique différentielle et traduction. Meta XVI (Actes du Colloque International de Linguistique et de Traduction) 1-2:17-24.

23. _____ 1964. Sémantique différentielle. Revue canadienne de Linguistique, X:7-14.

24. DEAL, L.V., 1970. Review: One-syllable Words. by MOSER, H. Journal of Speech and Hearing Disorders, XXXV,3:290.

25. DELATTRE, P., 1969. An acoustic and articulatory study of vowel reduction in four languages. International Review of Applied Linguistics in Language Teaching, VII, 4: 295-325.

26. DI PIETRO, Robert J., 1971. Language Structures in Contrast. Rowley, Mass.: Newbury House.

27. DULSEY, Bernard, 1956. Gender Differences in Romance Language Cognates. Hispania, XXXIX:466-467.

28. DYEN, Isidore, 1962. The Lexicostatistical Classification of Malayopolynesian Languages. Language, XXXVIII:38-46.

29. _____ 1963. Lexicostatistically determined Borrowing and Taboo. Language, XXXIX:60-66.

30. _____ 1964. On the Validity of Comparative Lexicostatistics. Proceedings of the Ninth International Congress of Linguists. Den Haag/Paris: Mouton, pp. 238-250.

31. _____ JAMES, A.T. and COLE, J.W.L., 1967. Language Divergence and Estimated Word Retention Rate. Language, XLIII: 150-71.

32. ELLEGARD, Alvar, 1959. Statistical Measurement of Linguistic Relationship. Language, XXXV:131-156.

33. ELLIS, Jeffrey, 1966. Towards a General Comparative Linguistics. Den Haag/Paris: Mouton.

34. FASOLD, Ralph W., 1970. Two Models of Socially Significant Linguistic Variation. Language, XLVI:551-563.

35. GAMMON, E., Quantitative Linguistic Typologies. Actes du 10ème Congrès International des Linguistes (Vol. III). Bucarest: Editions de l'ARSR, pp. 598-604.

36. GLEASON, H., 1968. Contrastive Analysis and Discourse Structure. Monograph Series on Language and Linguistics, Georgetown University, 21:39-63.

37. GOODENOUGH, Ward H., 1956. Componential Analysis and the Study of Meaning. Language, XXXII:195-216.

38. GREENBERG, Joseph H., 1966 . Language Universals. Current Trends in Linguistics (Vol. III) SEBEOK, T., ed. Den Haag/ Paris: Mouton, pp. 61-112.

39. _____ 1956. The Measure of Linguistic Diversity. Language, XXXII:109-115.

40. _____ 1959. A Method for Measuring Functional Yield as Applied to Tone in African Languages. Monograph Series on Languages and Linguistics, Georgetown University, XII:7-16.

41. GREENBERG, Joseph H., 1957. The Nature and Uses of Linguistic Typologies. International Journal of American Linguistics, XXIII,2.

42. _____ 1954. A Quantitative Approach to the Morphological Typology of Language. Method and Perspective in Anthropology. SPENCER, R.F., ed. Minneapolis: Minnesota University Press.

43. _____ 1960. A Quantitative Approach to the Morphological Typology of Language. International Journal of American Linguistics, XXVI:3.

44. _____ 1966. Synchronic and Diachronic Universals in Phonology. Language, XLII:508-517.

45. _____ ed., 1963. Universals of Language. Cambridge: M.I.T. Press.

46. GRIGORIEV, V.P., 1966. O neotorix voprosex interlingvistiki. Voprosy Jazykoznenija, III:1-15.

47. GRIMES, Joseph E. and AGARD, Frederick B., 1959. Linguistic Divergence in Romance. Language, XXXV:598-604.

48. GRIMES, J.E., 1964. Measures of Linguistic Divergences. Proceedings of the Ninth International Congress of Linguists. Den Haag/Paris: Mouton, pp. 44-50.

49. GUITER, H., 1969. Correlations de signifiants et de signifies dans les langues romanes. Travaux de Linguistique et de Littérature, VII, 1:131-159.

50. GUMPERZ, John J., 1958. Phonological Differences in three Hindi Dialects. Language, XXXIV:212-224.

51. HAMMER, J. and RICE, F., 1965. A Bibliography of Contrastive Linguistics. Washington: Center for Applied Linguistics.

52. HARARY, Frank and PAPER, Herbert H., 1957. Toward a General Calculus of Phonemic Distribution. Language, XXXIII: 143-169.

53. HARRIS, Zelig, 1968. Mathematical Structures of Language. New York: Interscience.

54. HARWOOD, F.W., 1955. Axiomatic Syntax: The Construction and Evaluation of a Syntactic Calculus. Language, XXXI: 409-413.

55. HARWOOD, F.W. and WRIGHT, Alison M., 1956. Statistical Study of English Word Formation. Language, XXXII:260-273.

56. HAUGEN, Einar, 1964. Review: Universals of Language. by GREENBERG, J.H., ed. Language, XL:260-269.

57. HERDAN, Gustav, 1964. Mathematics of Genealogical Relationships between Languages. Proceedings of the Ninth International Congress of Linguists. Den Haag/Paris: Mouton, pp. 51-60.

58. HOCKETT, G.F., 1967. Language, Mathematics and Linguistics. Den Haag/Paris: Mouton.

59. _____ 1967. The Quantification of Functional Load. Word, XXIII, 1-3:300-320.

60. HORALEK, K., 1967. Semantika textu e hlediska prekladatelskeho. Slovo, XXVIII, 2:113-120.

61. HORNE, K.M., 1966. Language Typology. 19th and 20th Century Views. Washington: Georgetown University 10 LAL.

62. ILIESCU, M., 1969. Reassemblances et dissemblances entre les langues romanes du point de vue de la morpho-syntaxe verbale. Revue de Linguistique Romane, XXXIII, 129-130:113-132.

63. IVIC, P., 1962. On the Structure of Dialect Differentation. Word, XVIII, 53 pp.

64. _____ 1964. Structure and Typology of Dialectal Differentiation. Proceedings of the Ninth International Congress of Linguists. Den Haag/Paris: Mouton, pp. 115-130.

65. JAKOBSON, R., 1962. K xarakteristike jevrazijskogo jazykovogo sojuza. Selected Writings (Vol. I) Paris: Mouton, pp. 117-136.

66. _____ 1962. Sur la théorie des affinités phonologiques entre les langues. Selected Writings (Vol. I) Paris: Mouton, pp. 234-246.

67. _____ FANT, C.G. and HALLE, M., 1967. Preliminaries to Speech Analysis. Cambridge: M.I.T. Press.

68. JENSEN, M.K., 1966. L'idée de ressemblance phonetique. Cahiers de Linguistique Theoretique et Appliquée, 3:91-97.

69. JESPERSEN, Otto, 1912. Growth and Structure of the English Language. Leipzig: Teubner (8th edition)/Oxford: Blackwell, 1935.

70. _____ 1924. A New Science: Interlinguistics. Oslo: Aschehoug.

71. KAROLY, S., 1967. Review: Sravnitelno-semasiologicheskie issledovaniya: Romanskie yasyki. by BUDAGOV, R.A. Altalanos Nyelveszeti Tanulmanyok, V:313-319.

72. KHUBCHANDANI, L.M. and GLOVER, W.W., 1969. An Application of Computer Techniques to Analysis of the Verb Phrase in Hindi and English. International Conference on Computational Linguistics (Abstracts). Stockholm: KVAL.

73. KIEFER, F., 1968. Mathematical Linguistics in Eastern Europe. New York: American Elsevier.

74. KING, R.D., 1967. Functional Load and Sound Change. Language, XLIII:831-852.

75. KIS, E., 1970. The Variability of Typological Features in the Evolution of the Romanian Language. Actes du dixième Congrès International des Linguistes. (Vol. III) Bucarest: Editions de l'ARSR, pp. 513-519.

76. KRAMSKY, Jiri. 1959. A Quantitative Typology of Languages. Language and Speech, II:75-85.

77. KRUPA, V., 1969. Review: Konferentsiya po problemam izucheniya universalnykh svoistv yazykov: tezisy dokladov. Jazykovedny Casopis, XX, 2:195-198.

78. KUCERA, Henry, 1963. Entropy, Redundancy and Functional Load in Russian and Czech. American Contributions to the Fifth International Congress of Slavists (Vol. I) Den Haag/Paris: Mouton, pp. 191-218.

79. _____ 1964. Statistical Determination of Isotopy. Prodeedings of the Ninth International Congress of Linguists. Den Haag/Paris: Mouton, pp. 713-722.

80. _____ and MONROE, G.K., 1968. A Comparative Quantitative Phonology of Russian, Czech and German. New York: American Elsevier.

81. LADEFOGED, Peter, 1969. The Measurement of Phonetic Similarity. International Conference on Computational Linguistics (Preprint 57) Stockholm: KVAL.

82. LEHMANN, W.P., 1962. Historical Linguistics. New York: American Elsevier.

83. LEJEUNE, Michel, ed., 1949. Actes du sixième Congrès International des Linguistes. Paris: Klincksieck, pp. 93-112, 409-416, 585-600.

84. LEVENSTON, E.A., 1966. A Classification of Language Differences. International Review of Applied Linguistics, IV, 3: 199-207.

85. LIEBERSON, Stanley, 1964. An Extension of Greenberg's Linguistic Diversity Measures. Language, XL:526-531.

86. LIEBOLD, H., 1969. Redundanz als sprachtypologisches Phaenomen. Fremdsprachen, III, 3:203-207.

87. LOTZ, J., 1967. Numerical Properties of Linguistic Structures. Anthropological Linguistics, IX, 4:1-4.

88. LUDVIKOVA, M., 1970. Symposium o algebraicke lingvistice konane ve Smolenicich. Slovo, XXXI, 3:287-288.

89. MACKEY, William F., 1967. Language Teaching Analysis. London: Longman, 1965/Bloomington: Indiana University Press.

90. _____ 1969. La rentabilité des minilangues. Cahiers Ferdinand de Saussure, XXVI (Mélanges Henri Frei) pp. 49-64.

91. _____ 1970a. Interference, Integration and the Synchronic Fallacy. Monograph Series on Languages and Linguistics, Georgetown University, 23:195-277.

92. _____ 1970b. An Algebra of Language Differences. Actes du dixième Congrès International des Linguistes. (Vol.III) Bucarest: Editions de l'ARSR, pp. 611-621.

93. _____ 1972a. Concept Categories as Measures of Cultural Differences. Man, Language and Culture: Studies in Honour of Claude Lévis-Strauss. Paris: Mouton (in the press).

94. _____ 1972b. Dialinguistic Identification. Studies in Honor of Einar Haugen. Cambridge: Harvard University Press (in the press).

95. _____ & LAFORGE, Lorne, 1972. Principes de didactique analytique. Paris: Didier.

96. MACKEY, W.F., SAVARD, J.-G. & ARDOUIN, P., 1971. Le vocabulaire disponible du français; Vol. I - Le vocabulaire concret usuel des enfants français et acadiens. Montréal/Paris: Didier.

97. MANCZAK, Witold, 1969. Le développement phonétique des langues romanes et la fréquence. Cracovie: Université Jagellon.

98. MANDELBROT, Benoit, 1961. On the Theory of Word Frequencies and on Related Markovian Models of Discourse. Structure of Language and its Mathematical Aspects. Providence, pp. 190-219.

99. MARTINET, André, 1949. Interlinguistique: Rapport préliminaire. Actes du sixième Congrès International des Linguistes. Paris: Klincksieck, pp. 93-112.

100. _____ 1967. Réflexions sur les universaux du langage. Folia Linguistica, I, 3-4:125-134.

101. _____ et al., 1956. Are there areas of affinité grammaticale as well as of affinité phonologique cutting across genetic language families? Proceedings of the Seventh International Congress of Linguists. London: International University Booksellers, pp. 439-459.

102. MESHCHANINOV, I.I., 1967. Primykanije v razlicnykh sintaksiceskikh sistemakh. Voprosy Jazykoznanija, III:3-17.

103. MEYERSTEIN, R.S., 1970. Functional Load. Den Haag/Paris: Mouton.

104. MILEWSKI, T., 1970. Voraussetzungen einer typologischen Sprachwissenschaft. Linguistica, LIX:62-107.

105. MOSSNER, F., 1969. Welchen Einblick gibt uns die Faktoranalyse in den Stasbau der natuerlichen Sprachen? Zeitschrift fuer Phonetik, Sprachwissenschaft und Kommunicationsforschung, XXII, 3:243-254.

106. MOUNIN, Georges, 1963. Les problèmes théoriques de la traduction. Paris.

107. NIDA, Eugene A., 1964. Toward a Science of Translation. Leiden: Brill.

108. PACIFIC CONFERENCE ON CONTRASTIVE LINGUISTICS AND LANGUAGE UNIVERSALS. Proceedings. Honolulu: University of Hawaii Press, 1971.

109. PANFILOV, V.Z., 1969. O zadachakh tipologicheskikh issledovanii i kriterii ikh tipologicheskoi klassifikatsii yazykov. Voprosy Iazykoznaniia, IV:3-15.

110. PAPP, F., 1966. Mathematical Linguistics in the Soviet Union. Den Haag/Paris: Mouton.

111. PARKER-RHODES, A.F., 1964. Is there an Interlingual Element in Syntax? Proceedings of the Ninth International Congress of Linguists. Den Haag/Paris: Mouton, pp. 176-190.

112. PETROVICI, Emil, 1969. Interpénétration des systèmes linguistiques. Actes du dixième Congrès International des Linguistes. (Vol. I) Bucarest: Editions de l'ARSR, pp. 37-74.

113. PIERCE, J.E., 1962. Possible Electronic Computation of Typological Indices for Linguistic Structures. International Journal of American Linguistics, XXVIII, pp. 215-226.

114. _____ 1957. A Statistical Study of Consonants in New World Languages. International Journal of American Linguistics, XXIII:36-45, 94-108.

115. PJOTROVSKIJ, R.G., 1969. Entropy and Redundancy in Four European Languages. Statistical Methods in Linguistics, V:34-35.

116. _____ Informacionnyje izmerenija jazyka. Leningrad: 1969.

117. PLATH, W., 1963. Mathematical Linguistics. Trends in European and American Linguistics 1930-1960. Mohrmann, Sommerfelt and Whatmough, eds. Utrecht-Antwerp: Spectrum.

118. REED, D.W. & SPICER, J.L., 1952. Correlation Methods of Comparing Idiolects in a Transition Area. Language, XXVIII:231-36.

119. REVZIN, I.I., 1967. Method modelirovanije i tipologija slavjanskix jazykov. Moscow.

120. _____ 1962. Modeli Jazyka. Moscow.

121. RHEINFELDER, H., 1970. Review: Sprachen vergleichbar und unvergleichlich. by WANDRUSZA, M. Die neueren Sprachen, LXIX, 5: 250-253.

122. SANKARAN, C.R., 1969. A Note on the Quantitative and Qualitative Definitions of the Alpha-phonoid. Phonetica, XX, 2-4: 74-80.

123. SANKOFF, D., 1970. On the Rate of Replacement of Word-meaning Relationships. Language, XLVI, 3:564-569.

124. SAPORTA, S., 1957. Methodological Considerations Regarding a Statistical Approach to Typologies. International Journal of American Linguistics, XXIII:109-113.

125. SELINKER, Larry, 1966. A Psycholinguistic Study of Language Transfer. Washington, D.C., Georgetown University, Ph.D. Dissertation.

126. SKALMOWSKI, W. & VAN OVERBEKE, M., 1969. Computational Analysis of Interference Phenomena on the Lexical Level. Tijdschrift van het Instituut voor Toegepaste Linguistiek, pp. 92-193.

127. _____ & VAN OVERBEKE, M., 1969. Computational Analysis of Interference Phenomena on the Lexical Level. International Conference on Computational Linguistics (Preprint 16) Stockholm: KVAL.

128. SPANG-HANSSEN, Henning, 1959. Probability and Structural Classification in Language Description. Copenhagen: Rosenkilde and Bagger.

129. TEETER, Karl V., 1963. Lexicostatistics and genetic relationship. Language, XXXIX, pp. 638-648.

130. USPENSKIJ, B.A., 1962. Principy Strukturnoi Tipologii. Moscow.

131. VOEGELIN, C.F., VOEGELIN, F.M., WURM, S., O'GRADY & MATSUDA, T., 1963. Obtaining an Index of Phonological differentiation from the construction of non-existent Minimax systems. International Journal of American Linguistics, XXIX, 1, pp. 5-28.

132. VOGT, H., 1969. Review: "Britain-USA: A survey in key-words" by: I. Friebel & H. Haendel, eds. Die Neueren Sprachen, LXVIII, 9, pp. 463-466.

133. WANDRUSZKA, Mario, 1971. Interlinguistik: Umrisse einer neuen Sprachwissenschaft. Munich: Piper.

134. _____ 1969. Sprachen, vergleichbar und unvergleichlich. Munich: Piper.

135. WINTER, Werner, 1970. Some Basic Difficulties in the Application of Quantifying Techniques to Morphological Typology. Actes du dixième Congrès International des Linguistes, (Vol. III). Bucuresti: Editions de l'ARCS, pp. 545-549.

136. ZIPF, G.K., 1935. The Psycho-Biology of Language. An Introduction to dynamic philology. Boston: Houghton Mifflin.